# Studying the Influence of Metal Ions on 5' Splice Site Formation in the Group II Intron *Sc*.ai5$\gamma$ by Single-Molecule Fluorescence

DISSERTATION

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Sebastian König

aus

Deutschland

Promotionskomitee:

Prof. Dr. Roland K.O. Sigel (Vorsitz und Leiter der Dissertation)

Prof. Dr. David Rueda

Prof. Dr. Benjamin Schuler

Zürich, 2014

*Gewidmet meinen Eltern, Norbert und Rosemarie König.*

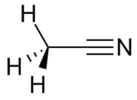
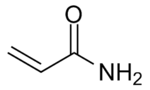
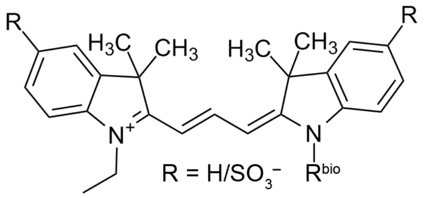
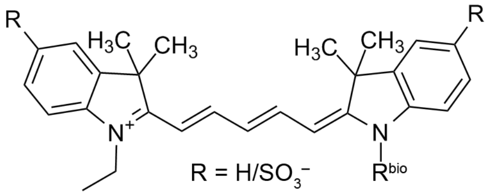
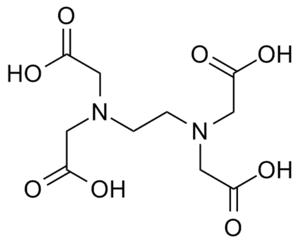
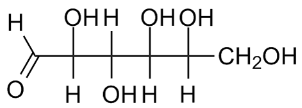
*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.*

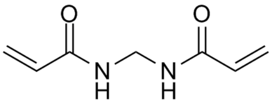
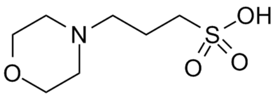
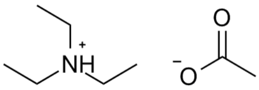
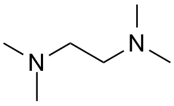
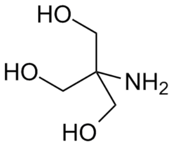
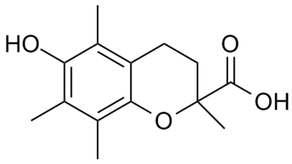
\- James D. Watson and Francis H. Crick [1].

# Contents

# Chapter 1

# List of (organic) chemicals used

| Name | Chemical structure/explanation | Supplier |
|---|---|---|
| Acetonitrile |  | Sigma-Aldrich<br>Buchs,<br>Switzerland |
| Acrylamide |  | National diagnostics,<br>Atlanta,<br>Georgia, USA |
| Cy3 |  | GE Healthcare,<br>Glattbrugg,<br>Switzerland |
| Cy5 |  | GE Healthcare,<br>Glattbrugg,<br>Switzerland |
| EDTA |  | Sigma-Aldrich,<br>Buchs,<br>Switzerland |
| D-Glucose |  | Sigma-Aldrich,<br>Buchs,<br>Switzerland |

| | | |
|---|---|---|
| N,N'-Methylene-bisacrylamide |  | National diagnostics,<br>Atlanta,<br>Georgia, USA |
| MOPS |  | Sigma-Aldrich,<br>Buchs,<br>Switzerland |
| TBE | Tris/Borate/EDTA | - |
| TEAA |  | Sigma-Aldrich,<br>Buchs,<br>Switzerland |
| TEMED |  | Sigma-Aldrich,<br>Buchs,<br>Switzerland |
| Tris |  | Sigma-Aldrich,<br>Buchs,<br>Switzerland |
| Trolox ® |  | Sigma-Aldrich,<br>Buchs,<br>Switzerland |

# Chapter 2

# Abbreviations and physical constants (alphabetical order)

**A:** adenine
**ADP:** adenosine-5’-diphosphate
**AFM:** atomic force microscopy
**AIC:** Akaike information criterion
**ALEX:** alternating laser excitation
**AMPPNP:** 5’-adenylyl-$\beta$-$\gamma$-imidodiphosphate
**ANOVA:** analysis of variance
**APD:** avalanche photodiode
**aRNA:** antisense ribonucleic acid
**ATP:** adenosine-5’-triphosphate
$\beta$**:** stretching exponent
**BIC:** Bayesian information criterion
**BOBA FRET:** bootstrap-based analysis of smFRET data, name of a software
**BODIPY:** boron-dipyrrolmethene
**bp:** base pair
$c$**:** concentration
**C:** cytosine
**CARD:** caspase activation and recruitment domain
**CCD:** charge-coupled device
**CSSR:** name of a software to perform hidden Markov modelling
**cumP:** cumulative probability
**Cy3/Cy5:** cyanine dye 3/5
**ddNTP:** dideoxynucleoside triphosphate
**Ded1:** ATP-dependent RNA helicase DED1
**DNA:** deoxyribonucleic acid
**dsRNA:** double-stranded RNA

**$\varepsilon$:** extinction coefficient
**EBS:** exon-binding site
**eIF4A:** eukaryotic initiation factor 4A
**EM-CCD:** electron-multiplying charge-coupled device
**extIWS:** extended Irving-Williams series
**FACS:** fluorescence-activated cell sorting
**FCS:** fluorescence correlation spectroscopy
**FISH:** fluorescence *in situ* hybridisation
**FRET:** Förster resonance energy transfer
**G:** guanine
**$\Gamma$:** gamma distribution
**GFP:** green fluorescent protein
**GOF:** goodness of fit
**h:** Planck constant, 6.626 $10^{-34}$ J s
**$\eta$:** refractive index
**HaMMy:** name of a software to perform hidden Markov modelling
**HCV:** hepatitis C virus
**HILO:** highly inclined and laminated optical sheet
**HMM:** hidden Markov model/modelling
**HPLC:** high performance liquid chromatography
**$I$:** (fluorescence emission) intensity
**IBS:** intron-binding site
**$J(\lambda)$:** overlap integral
**$K$:** Stern-Vollmer constant for collisional quenching
**$K_A$:** association constant
**$K_D$:** dissociation constant
**$K_m$:** Michaelis-Menten constant
**$\kappa^2$:** Förster resonance energy transfer orientation factor
**$k_{nr}$:** kinetic rate constant of non-radiative decay
**$k_r$:** kinetic rate constant of radiative decay
**$\lambda$:** wavelength
**LCM:** laser confocal microscopy
**M, M(n), $M^{n+}$:** (n-valent) metal ion
**MANOVA:** multivariate analysis of variance
**ME:** maximum evidence
**miRNA:** micro ribonucleic acid
**ML:** maximum likelihood
**mRNA:** messenger ribonucleic acid
**Mss116:** ATP-dependent RNA helicase MSS116, mitochondrial

**$\nu$:** frequency
**ncRNA:** non-coding ribonucleic acid
**NMR:** nuclear magnetic resonance
**NPH-II:** nucleoside triphosphate phosphohydrolase II
**NS3:** non-structural protein 3
**NSOM:** near-field scanning microscopy
**NTP:** nucleoside triphosphate
**P:** fluorescence polarisation
**P-value:** strength of evidence against the null hypothesis $H_0$
**PAMP:** pathogen-associated molecular pattern
**PC:** photon counts
**PCR:** polymerase chain reaction
**PEG:** polyethylene glycol
**PMT:** photomultiplier tube
**PIFE:** protein-induced fluorescence enhancement
**PSF:** point-spread function
**qPCR:** quantitative polymerase chain reaction
**QuB:** name of a software
**$QY$:** quantum yield
**$r$:** fluorescence anisotropy
**$R$:** distance; Pearson product-moment correlation coefficient
**$R_0$:** Förster radius
**RD:** regulatory domain
**RIG-I:** retinoid acid inducible-gene protein I
**RNA:** ribonucleic acid
**RNase P:** ribonuclease P
**rRNA:** ribosomal ribonucleic acid
**$\sigma$:** standard deviation
**SF:** (helicase) superfamiliy
**SIRA:** sif reader and analyser
**siRNA:** small interfering RNA
**SMART:** single molecule analysis research tool, name of a software
**smFRET:** single-molecule Förster resonance energy transfer
**snoRNA:** small nucleolar ribonucleic acid
**SNR:** signal-to-noise ratio
**snRNA:** small nuclear ribonucleic acid
**ssDNA:** single-stranded deoxyribonucleic acid
**ssRNA:** single-stranded ribonucleic acid
**$\tau$:** exponential decay constant

---

**TCSPC:** time-correlated single photon counting
**TDP:** transition density plot
**TIRF(M):** total internal reflection fluorescence (microscopy)
**TPM:** tethered-particle microscopy
**tRNA:** transfer ribonucleic acid
**U:** uracil
**UTR:** untranslated region
**vbFRET:** variational bayesian inference for smFRET time series, name of a software
**YxiN:** ATP-dependent RNA helicase YxiN

# Chapter 3

# Introduction

## 3.1 RNA folding, function and the importance of metal ion cofactors

Ribonucleic acids (RNA) are essential for (de-)coding, regulation and expression of genetic information. Due to their polyanionic character, formation of structure and hence functionality is inextricably linked to the presence of metal ions. This section surveys the building blocks of RNA and their cation binding properties, as well as general characteristics of RNA folding. Subsequently, RNA functions in the living cell are summarised with a particular focus on RNA enzymes (*ribozymes*) and how their folding and function is regulated through the action of metal ions.

### 3.1.1 RNA building blocks and synthesis

RNA consists of three basic building blocks: (i) cyclic riboses, (ii) phosphodiesters, and (iii) heterocyclic bases [2]. In an RNA molecule, sugar rings, interconnected via phosphodiester linkages, form the so-called backbone, while the purines guanine (G) and adenine (A), as well as the pyrimidines cytosine (C) and uracil (U) are attached to the sugar rings (Figure 3.1) [3].

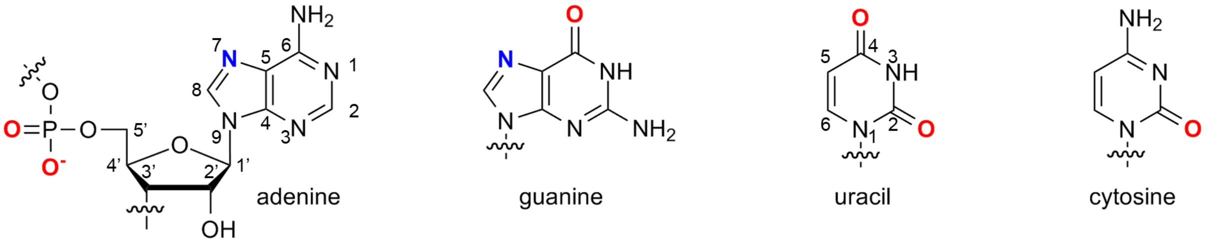


Figure 3.1: Structure of the RNA phospho-sugar backbone and the four most abundant ribonucleobases. The numbering scheme is given to indicate the position of functional groups in purine and pyrimidine bases. Major metal ion binding sites are highlighted in red and blue. Figure adapted from [2].

*In vivo*, RNA is synthesised from a DNA template and nucleoside 5'-triphosphate molecules (NTPs) by a large multidomain enzyme called RNA polymerase (RNAP), which catalyses the $Mg^{2+}$-dependent nucleophilic attack of the 3'OH on the $\alpha$-phosphate of the dNTP (Figure 3.2A) [4]. This mechanism implies that RNA molecules display directionality, *i.e.* a phosphate group is attached to one end of the RNA (5'-end) and a hydroxyl group at the other (3'-end). In contrast to RNAP-mediated RNA polymerisation, chemical synthesis of oligonucleotides is typically performed from the 3'-end to the 5'-end [5]. For example, the RNA sequences used in this work were synthesised chemically using the phosphoramidite approach illustrated in Figure 3.2B [5].

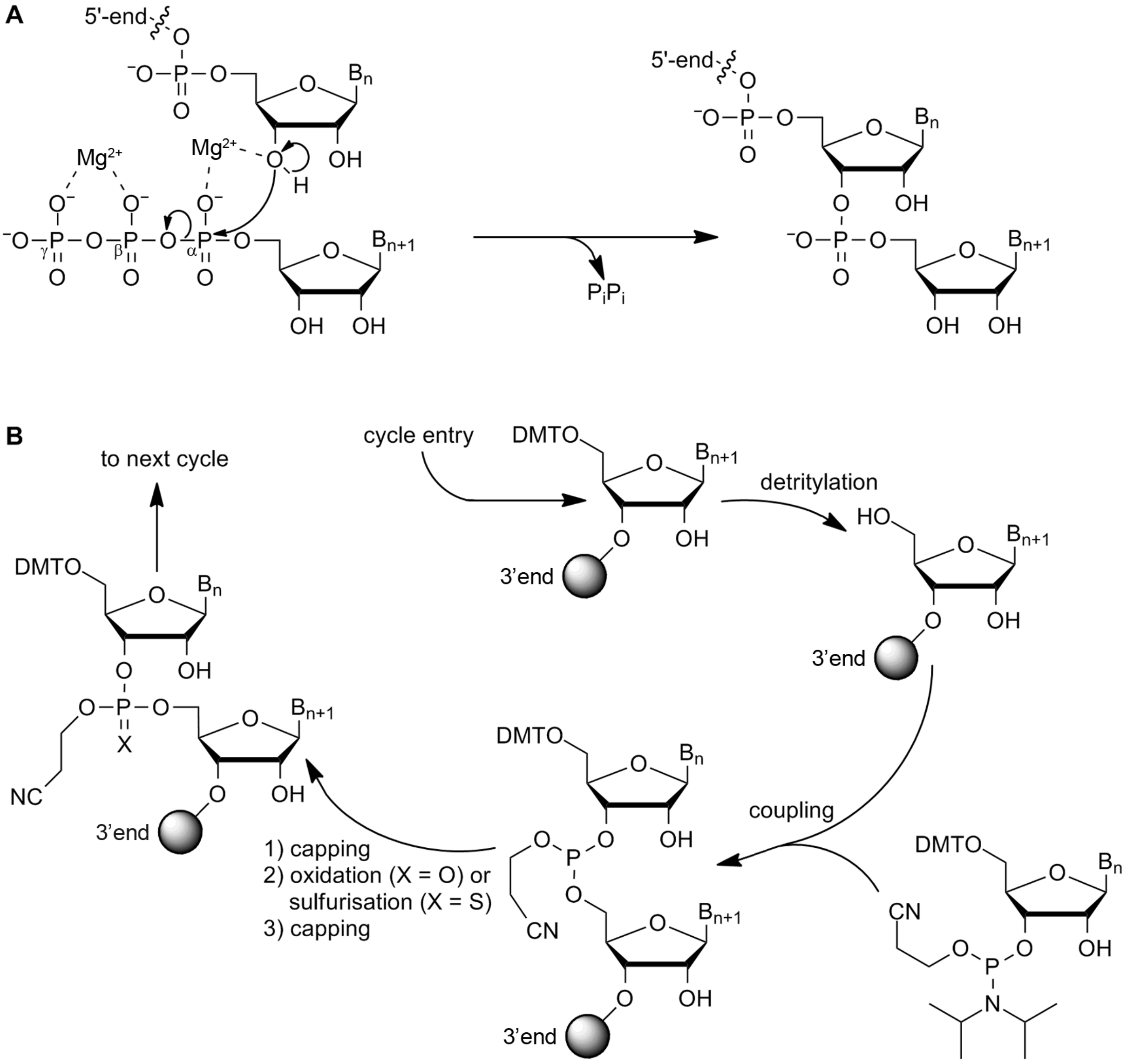


Figure 3.2: Biological and chemical RNA synthesis. (A) RNAP polymerises dNTPs in a $Mg^{2+}$-dependent manner from the 5'-end to the 3'-end. (B) Phosphoramidite-based chemical synthesis occurs in the 3'->5' direction. B = base, sphere = solid support, DMT = dimethoxytrityl. Adapted from [4, 5].

### 3.1.2 Base-pairing and RNA secondary structure

The purine and pyrimidine bases attached to RNA are able to form specific hydrogen bonds that depend on geometrical correspondence of hydrogen bond donors and acceptors. The most relevant base-pairing scheme in a biological context is called Watson-Crick base-pairing (also: *canonical base-pairing*). As depicted in Figure 3.3A, Watson-Crick base pairing leads to specific interaction of guanine with cytosine via three hydrogen bonds, while adenine and uracil form two hydrogen bonds [3]. It should be noted that other base-pairing schemes are also known to be formed by RNA nucleobases, for example so-called G-U wobbles and Hoogsteen base-pairs, albeit they are less frequently encountered in a biological context [8, 9, 10, 11].

Secondary structure refers to a (local) regular three-dimensional arrangement of a biopolymer. In RNA, the predominant secondary structure element is the A-form double helix, in which two phospho-sugar backbones are coiled around the same axis in a right-randed fashion and a contiguous run of Watson-Crick base-paired nucleobases is sequestered inside (Figure 3.3B) [1, 12]. In contrast to the B-form helix that is preferably adopted by DNA, sugar puckers in the A-form helix adopt the C3'-endo instead of the C2'-endo conformation (Figure 3.3C) [7]. Since steric reasons prevent the ribose units in RNA from adopting the C2'-endo conformation, RNA is not known to adopt the B-form at all [13], albeit other secondary structure elements have been reported, for example Z-RNA, G-quadruplexes and *i*-motifs [14, 15, 16, 17]. In turn,

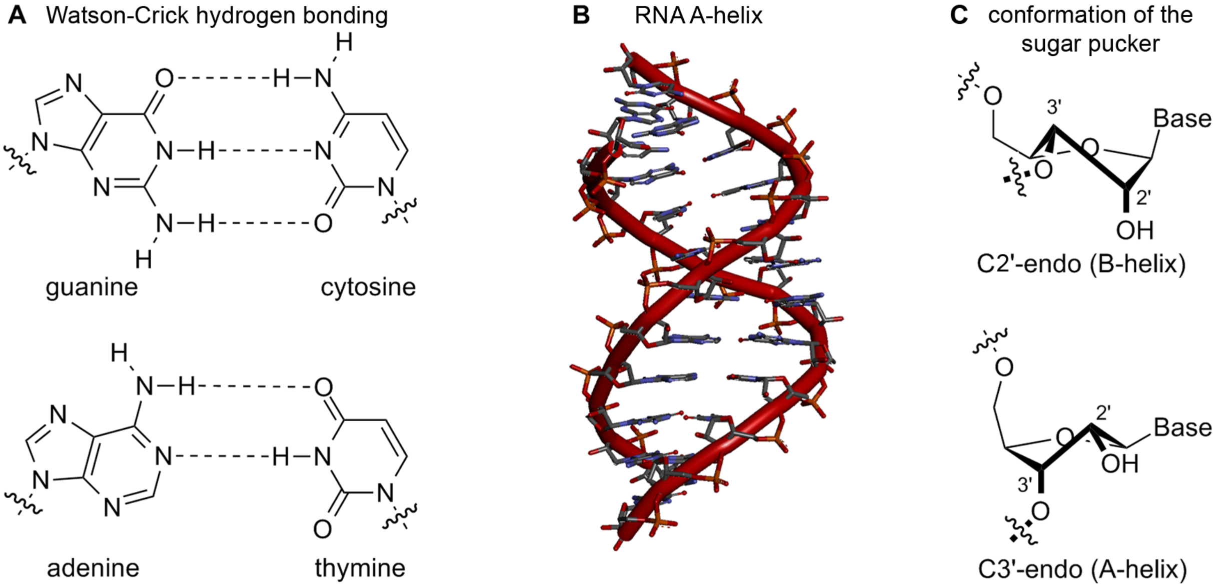


Figure 3.3: (A) Watson-Crick base-pairing. Guanine and cytosine form three hydrogen bonds, while adenine and thymine form two hydrogen bonds. (B) The A-form is the predominant secondary structural motif found in RNA. Drawn with Accelrys Discovery Studio client 3.1 (Accelrys, Inc., San Diego, CA, USA) from PDB entry 1LNT [6]. The conformations of the sugar pucker depends on the helical context. In RNA, a steric clash of the 2-OH with the 3'-OH precludes the formation of the B-form helix. Figure adapted from [7].

it should be noted that the A-form helix is by far the most important architecture *in vivo*.

### 3.1.3 RNA folding and tertiary structure

RNA folding is generally a hierarchical process [22]. During the first steps of RNA folding, secondary structure elements form from randomly coiled RNA, which predominantly consist of stretches of A-form RNA. At this stage, structural diversity is achieved by alternating stretches of double- and single-stranded RNA, yielding for example isolated duplexes, hairpins, bulges, internal loops and junctions (Figure 3.4) [23]. Typically, an RNA molecule adopts its global three-dimensional fold (*tertiary structure*) upon formation of long-range tertiary interactions between pre-organised secondary structure elements, albeit exceptions are known [24, 25]. Numerous tertiary interactions have been precedented, including coaxial stacking, adenosine

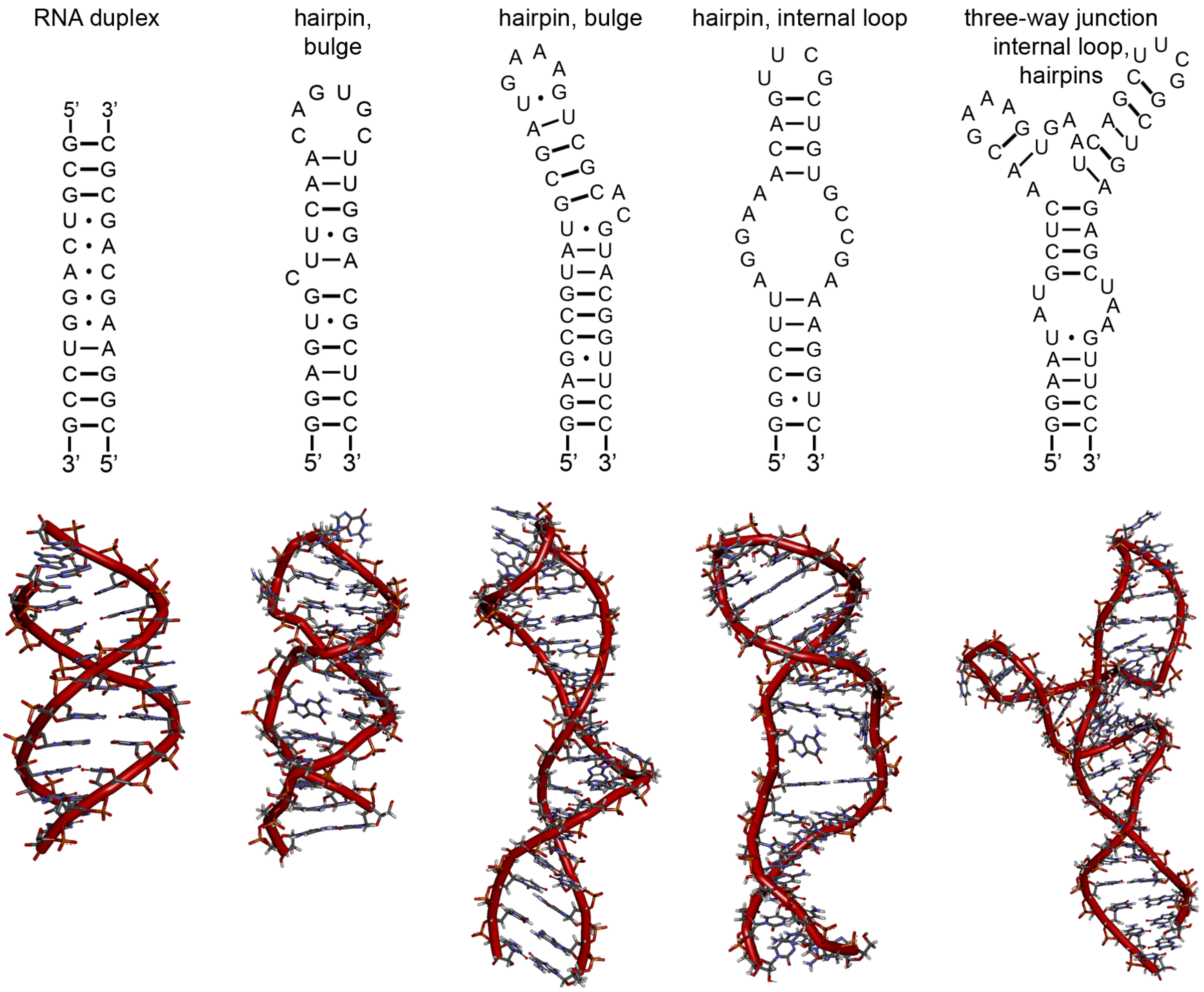


Figure 3.4: Different secondary structure elements are known to be formed by RNA, consisting of alternated arrangements of A-form helices and stretches of single-stranded RNA. Structures drawn with Accelrys Discovery Studio client 3.1 (Accelrys, Inc., San Diego, CA, USA) from PDB entries 1LNT [6], 1NBR [18], 1R2P [19], 2EUY [20], 2LU0 [21].

platforms, base triples and triplexes, tetraloop-tetraloop acceptor motifs, metal core motifs, loop-loop interactions, and psedoknots [23]. Interaction of a hairpin loop with its complementary single-stranded sequence is in the focus of the work presented herein.

### 3.1.4 RNA function, ribozymes and group II introns

RNA was originally thought to function solely as a messenger that is first transcribed from a DNA template (*gene*) and subsequently conveys the genetic information to the ribosome for protein synthesis to occur in a sequence-specific manner (*messenger RNA*, mRNA) [36]. However, it has become evident that almost the entire genome is transcribed in most eukaryotic cells, creating a vast amount of non-coding RNA (ncRNA), including ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), micro RNA (miRNA), small interfering RNA (siRNA), and riboswitches (Table 3.1) [26]. In general, these ncRNAs usually constitute structural scaffolds for proteins that function together with the RNA [35]. However, the discovery of group I introns and the RNA component of *E. coli* and *B. subtilis* RNase P demonstrated the existence of catalytically active RNAs (ribozymes) that carry out important functions in their own right [35, 37].

Table 3.1: Non-exhaustive Overview on non-coding RNAs and their functions in the living cell. List adapted from [26, 27, 28].

| Abbreviation | Meaning | Biological role(s) |
|---|---|---|
| rRNA | ribosomal RNA | integral component of the ribosome, which synthesises polypeptides across all living organisms [29]. |
| tRNA | transfer RNA | tRNAs are “adapter molecules” that transfer amino acids onto a nascent polypeptide chain [30]. |
| snRNA | small nuclear RNA | The spliceosome consists of snRNA and proteins and excises non-coding sequences from precursor mRNA (splicing) [31]. |
| snoRNA | small nucleolar RNA | snoRNA incorporate methyl groups and pseudouracyls in rRNAs, tRNAs, and snRNAs [32]. |
| miRNA | micro RNA | Endogenous miRNAs regulate protein expression of most genes in animals via the RNA-induced silencing complex [33]. |
| siRNA | small interfering RNA | siRNAs have similar function as miRNA, but they are exogenous [34]. |
| riboswitches | - | Riboswitches are located on mRNA and they undergo conformational changes in response to small metabolites leading to regulation of translation [35]. |

Naturally occurring ribozymes that do not depend on protein cofactors to achieve catalytic activity generally carry out phosphodiester chemistry and are classified as *cleaving* and *splicing* ribozymes depending on the type of reaction that is catalysed (Table 3.2) [38]. Despite their rather short length ($<$ 200 nucleotide units), *cis*-cleaving ribozymes adopt diverse three-dimensional folds and are involved in different aspects of cellular metabolism. Hammerhead, hairpin, hepatitis delta virus and Varkud satellite ribozymes are the product of a so-called rolling-circle replication, in which long multimeric RNAs are processed to form short monomers [35]. They mostly occur in satellite RNAs, *i.e.* stretches of viral single-stranded RNA that dependent on helper viruses for their replication [39]. The second subgroup includes functionally diverse ribozymes that are predominantly found on eukaryotic pre-mRNA or bacterial mRNA [35]. Ribonuclease P (RNase P) catalyses the maturation of the 5' end of tRNA by performing a *trans* cleavage reaction [40]. The structurally distinct group I and group II introns are large catalytic RNAs ($\leq$3000 nucleotides), which catalyse their own excision from pre-mRNA, followed by ligation of the adjacent sequences (*exons*), a reaction referred to as *cis*- or self-splicing [41]. Also, both group I and group II introns are known to constitute mobile genetic elements capable of reinserting in the genome [42].

Table 3.2: Overview on the most important classes of ribozymes, their main functions, size and distribution. List adapted from [35].

| Class | Members | Functions | Size (nucleotides) | Prevalence |
|---|---|---|---|---|
| *cis*-cleaving ribozymes | hammerhead ribozyme | processing of multimeric transcripts during replication | 65 | viroids, plant viral RNA, eukaryotes |
| | hairpin ribozyme | | 75 | plant viral satellite RNA |
| | Varkud satellite ribozyme | | 155 | *Neurospora* satellite RNA |
| | Hepatits delta virus ribozyme | | 85 | *Human* satellite virus |
| | CoTC ribozyme | transcription termination (?) | 190 | primates |
| | CPEB3 ribozyme | splicing regulation (?) | 70 | mammals |
| | glmS ribozyme | gene control | 170 | Gram-positive bacteria |
| *trans*-cleaving ribozymes | RNase P | tRNA processing | 140-500 | prokaryotes, eukaryotes |
| *cis*-cleaving group I introns | | self-splicing | 200-1500 | organelles, bacteria, bacteriophages, mitochondria |
| *cis*-cleaving group II introns | | self-splicing | 300-3000 | organelles, bacteria, archaea |

### 3.1.5 Group II intron structure and catalysis

Group II introns have not been identified in the nuclear genomes of eukaryotes, but they have received considerable attention in recent years because of their interesting evolutionary roles: There is accumulating evidence that group II introns are ancestors of the spliceosome, *i.e.* a large RNA-protein complex that excises performs splicing in eukaryotic cells, as well as retrotransposons, *i.e.* genetic elements that amplify themselves by means of an RNA intermediate [43, 44, 45]. Recent computational studies further motivate research on group II introns, as they reported that spliceosomal introns and retrotransposons comprise more than half of the human genome [45, 46]. Furthermore, they also provide a wealth of different folding motifs and long-range tertiary contact architectures that are interesting in their own right [19, 47, 48, 49, 50]. This section provides an account of group II intron structure and function.

Group II introns generally catalyse the nucleophilic attack of a hydroxyl group or a water molecule on an activated phosphodiester linkage, leading to the formation of a 3'-OH and a 5'-phosphate or a new phosphodiester linkage [51]. This chemical repertoire is used in three distinct reaction pathways: In the lariat pathway (Figure 3.5A), the 2'-OH of a specific adenosine residue (bulge adenosine) attacks the 5' splice site, followed by an attack of the 5'-exon 3'-OH on the 3' splice site [51]. Exons are joined together and the intron is released in its lariat form, which is reminiscent of the product of spliceosomal splicing. In the hydrolytic pathway (Figure 3.5B), the first step of splicing involves the nucleophilic attack of a water molecule, which leads to exon ligation and the release of a linear intron [52]. Group II introns can also be released as circular products (Figure 3.5C), a reaction that has been proposed to be initiated

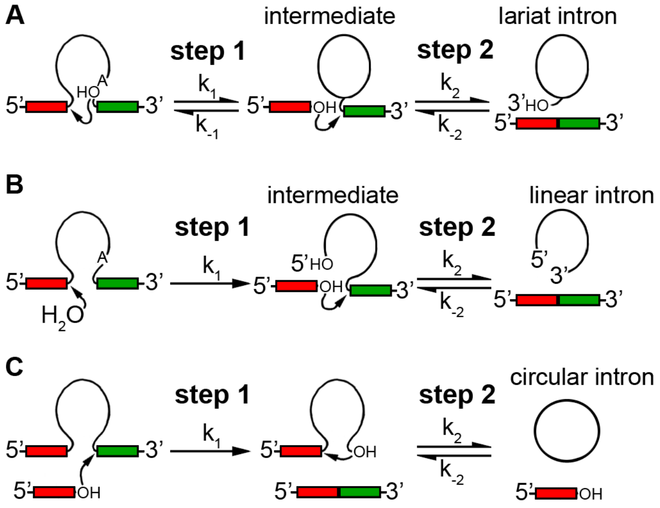


Figure 3.5: Different pathways of group II intron-mediated splicing and intron excision. (A) Analogous to spliceosomal splicing, the branching pathway release the intron in its lariate form. (B) The hydrolytic pathway leads to the formation of a linear splicing product. (C) Working model to explain intron circularisation. Figure adapted from [51].

by the nucleophlic attack of the 3'-OH of a free 5' exon [53, 54]. However, the lariat and the hydrolytic pathway are of greater relevance both *in vitro* and *in vivo* [55].

Group II introns owe their functional versatility to six subdomains, each of which plays a specific role in folding or catalysis (*D1-6*, Figure 3.6) [57]: Domain 1 is the scaffold for the assembly of the other domains and is consequently essential for splicing [58]. It also contains functionalities for exon and splice site recognition (*exon-binding sites*, EBS). Domain 2 has been proposed to serve as a conformational switch between the two steps of splicing, thereby

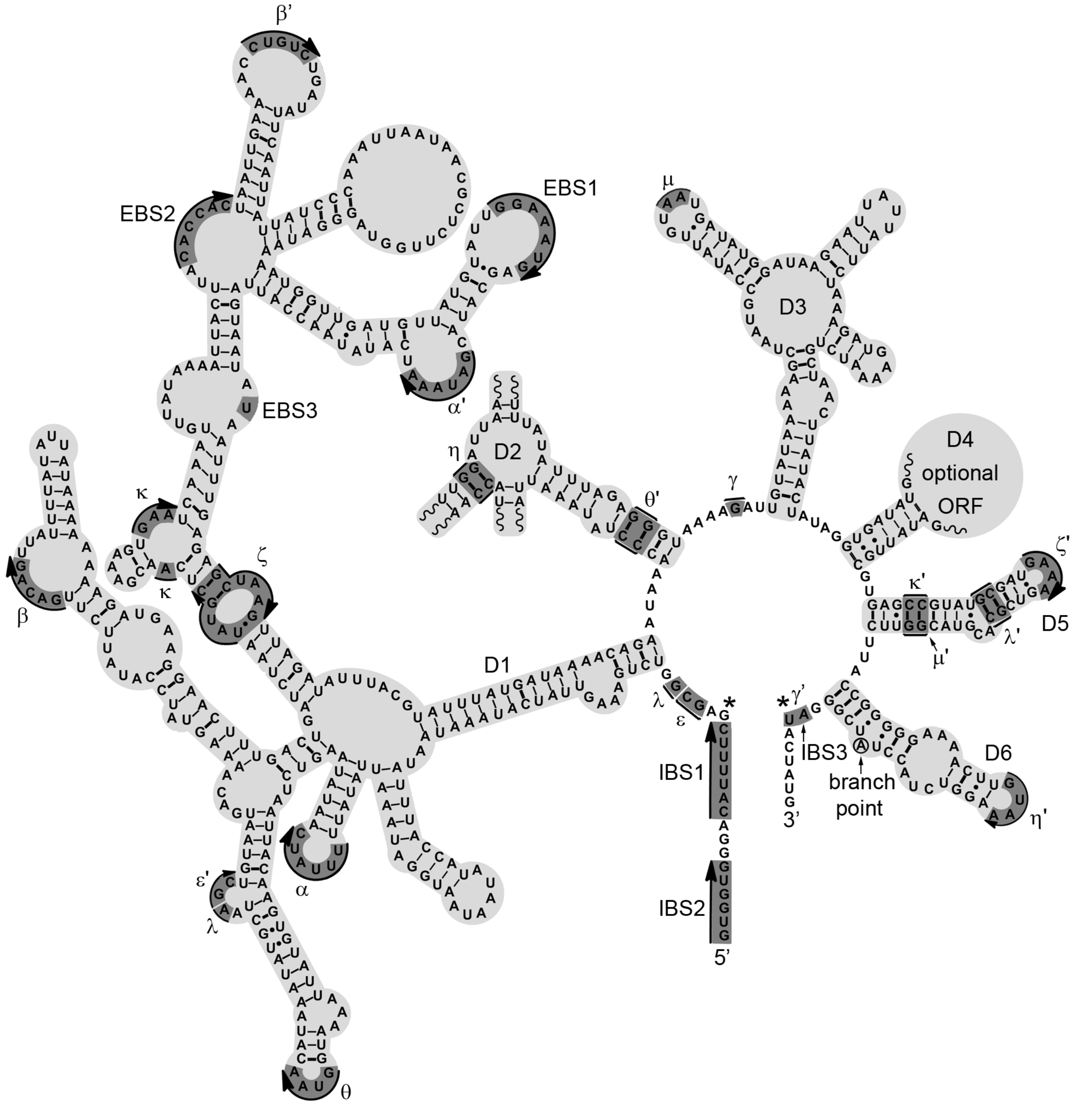


Figure 3.6: Secondary structure of the D135-L14 ribozyme derived from the Sc.ai5$\gamma$ group II intron from the *cox-1* gene in *S. cerevisiae*. Long-range interactions are highlighted in dark gray and named with Greek letters. Intron/exon binding sites are also shown (IBS1, IBS2 and EBS1, EBS2, highlighted in dark grey). The two splice sites are marked with an asterisk. Adapted from [56].

enhancing catalytic activity [21]. Domain 3 (*catalytic effector*) is not strictly required for catalysis, but its presence strongly enhances catalytic activity [59]. Domain 4 may contain an open reading frame that encodes a so-called maturase, a multi-domain protein that is thought to facilitate self-splicing and intron mobility *in vivo* [60]. Domain 5 (*catalytic domain*) is the most conserved part of the intron that contains numerous functionalities for D1 binding and is indispensable for catalytic activity [61, 62]. Finally, domain 6 contains the bulged adenosine, which is crucial for intron excision via the lariat-forming pathway (*vide supra*).

### 3.1.6 The importance of metal ions in RNA folding and catalysis

The polyanionic nature of the backbone entails that the formation of three-dimensional structure leads to an accumulation of negative charges in close space [2]. As a consequence, positively charged (metal) ions are required to overcome the electrostatic repulsion and to allow stable formation of an RNA secondary and tertiary structure [63]. Here, the oxygen atoms of the phosphodiester bridges located at the exterior of the double helix, especially the negatively charged non-bridging ones, constitute the major metal ion binding sites [64]. Further important cation interaction sites are provided via exocyclic nucleobase oxygens (purine O6, pyrimidine O2 and O4) and endocyclic nitrogens (purine N7) (Figure 3.1) [64]. Not surprisingly, the majority of RNA structures revealed metal ions interaction with the nucleic acid [65]. This section provides an overview of the importance of metal ions for RNA folding and catalysis.

Negative charges are usually screened by the monovalent alkali ions $Na^+$ and $K^+$, which are abundant at millimolar concentrations in the living cell (Table 3.3) [98, 99]. Specific roles of monovalent metal ions are nearly unprecedented, though, ribosomal rRNA has been reported to unfold upon depletion of $K^+$ and ribosomal activity is known to decrease in the presence of $Na^+$, which suggests that the role of the metal ion is unlikely to be limited to non-specific charge screening [67, 68, 69, 70, 71, 72]. Arguably, the most important divalent metal ion in RNA folding and catalysis is $Mg^{2+}$, which is due to its comparably high concentration and its strong propensity to coordinate to hard ligands like phosphates and exocyclic oxygens [66]. As a consequence, it is considered the natural cofactor for the majority of catalytically active RNA. For example, it has been shown to be directly involved in group II intron splicing [19, 48, 76, 77, 78, 79, 80], stabilisation of the ribosome [72, 81, 82], and regulation of $Mg^{2+}$ uptake through $Mg^{2+}$-sensing riboswitches [83, 84, 85, 86]. Even though other multivalent cations are known to play important roles in many aspects of cellular metabolism, often through direct coordination to proteins, they are often not considered in RNA folding and function, because their concentration is comparably low and they usually prefer nitrogen ligands over oxygen ligands [65, 100]. However, numerous *in vitro* studies have demonstrated that ribozyme activity can be tuned by the metal ion cofactor. For example, $Ca^{2+}$ inhibits group II intron-mediated splicing [78, 87], $Mn^{2+}$ strongly alters catalytic rates of the hairpin and the hammerhead ribozyme [88, 91, 101], $Zn^{2+}$ induces a sequential cleavage pattern in the hammerhead ribozyme [96], and the *Tetrahymena* group I intron retains its activity when

$Mg^{2+}$ is replaced by the organic polycation spermidine ($C_7H_{22}N_3^{3+}$) [97]. Most remarkably, $Fe^{2+}$ has recently been shown to confer on some RNAs a previously uncharacterised ability to catalyse single-electron transfer, expanding the chemical repertoire of naturally occurring ribozymes [93]. It has been speculated that in these cases, $Mg^{2+}$ is not the natural cofactor and/or that folding and activity of RNA can be controlled via a so-called metal ion switch,

Table 3.3: Concentration of various metal ions inside bacterial and mammalian cells, (putative) RNA interaction partners and biological role(s). N.D. = not determined. Concentrations taken from [66].

| Cation | Abundance, bacterial cells (mM) | Abundance, mammalian cells (mM) | (Putative) RNA interaction partner(s) and biological role(s) |
|---|---|---|---|
| $Na^+$ | N.D. | 10 | $Na^+$ inhibits ribosomal activity *in vitro* [67, 68, 69]. |
| $K^+$ | >10 | 140 | coordinates to all RNAs; rRNA unfolds in the absence of $K^+$ *in vitro* [70, 71, 72], strong stabilisation of RNA G-quadruplexes *in vitro* [73, 74, 75]. |
| $Mg^{2+}$ | > 10 | 30,<br>free: ∼1 | coordinates to all RNAs, group II intron folding/splicing [19, 48, 76, 77, 78, 79, 80] stabilisation of the ribosome [72, 81, 82] $Mg^{2+}$-sensing riboswitches [83, 84, 85, 86]. |
| $Ca^{2+}$ | 0.1 | 1 | $Ca^{2+}$ inhibits group II intron splicing *in vitro* [78, 87]. |
| $Mn^{2+}$ | 0.01 | N.D. | Replacement of $Mg^{2+}$ by $Mn^{2+}$ leads to decreased (hairpin rbz., [88]), comparable (RNase P, HDV rbz., [89, 90]), or increased activity (hammerhead rbz., [91]) *in vitro*. |
| $Fe^{2+}$ | 0.1 | N.D. | mediates contact between IRPs and the IRE on mRNA [92]; 23S rRNA, the P4-P6 domain and yeast $tRNA^{phe}$ catalyse electron transfer with $Fe^{2+}$ as cofactor *in vitro* [93]. |
| $Co^{2+}$ | < 0.01 | N.D. | cobalamin ($Co^{2+}$ chelated by a corrin ring) binds to $B_{12}$ riboswitches, regulation of vitamin $B_{12}$ synthesis *in vivo* [94, 95]. |
| $Ni^{2+}$ | < 0.01 | N.D. | N.D. |
| $Zn^{2+}$ | 0.1 | N.D. | induces sequential cleavage in hammerhead ribozymes *in vitro* [96]. |
| $C_7H_{22}N_3^{3+}$ | N.D. | N.D. | can replace $Mg^{2+}$ in *Tetrahymena* group I intron folding and catalysis *in vitro* [97] |

but systematic studies on cation-mediated RNA folding and catalysis are currently very scarce [101, 102, 103, 104].

In summary, stable formation of nucleic acid structures is not possible without metal ions and the majority of ions interact with RNA in a non-specifical manner [105]. In such an ion atmosphere, cations are in fast exchange and the distribution of positive charges is generally modelled using Poisson-Boltzmann theory [106]. Based on the results of X-ray diffraction studies, an estimated 10 % of metal ions are bound very specifically and locally [2, 76, 107]. It is generally believed that $Mg^{2+}$ is the most important cofactor, but the principles of metal ion selection by RNAs are still poorly understood [2]. Please refer to Table 3.3 for intracellular metal ion concentrations and an overview on cation-mediated RNA folding and catalysis.

## 3.2 Fluorescence and Förster resonance energy transfer

This section provides an account of the physical principles of fluorescence, including a photochemical differences between fluorescence and phosphorescence, definitions of the extinction coefficient, quantum yield, fluorescence lifetime, as well as fluorescence anisotropy. Moreover, different mechanism of fluorescence quenching are discussed, including Förster resonance energy transfer. Since fluorescence and Förster resonance energy transfer have nowadays found numerous practical applications, commonly used fluorophores and their use is biological research are also described.

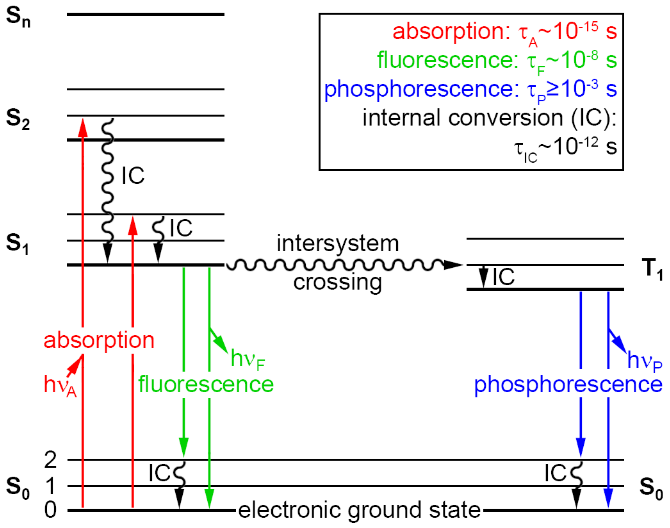


Figure 3.7: Jablonski diagram illustrating the photophysical processes underlying absorption, fluorescence and phosphorescence and their typical time constants.

### 3.2.1 Luminescence, fluorescence, and phosphorescence

Photoluminescence refers to the emission of light upon relaxation of an excited state electron, whereby the energy of the emitted photon corresponds to the energy difference between the excited and the ground state. As depicted in the Jablonski diagram (Figure 3.7), luminescence is generally preceded by the absorption of a photon (h$\nu$), which excites an electron in the ground state $S_0$ to a higher singlet quantum state ($S_1$, $S_2$, ..., $S_n$), a rapid process that takes place within $\tau_A \sim 10^{-15}$ s. Excitation is typically followed by rapid internal conversion (*IC* or *vibrational relaxation*) to the lowest vibrational energy level of the lowest excited state $S_1$ ($\tau \sim 10^{-12}$), a phenomenon that is commonly referred to as Kasha's rule [108].

Depending on the nature of the excited state prior to the radiative decay, one formally distinguishes fluorescence from phosphorescence. In fluorescence, the electron decays directly from $S_1$ to $S_0$, a process that occurs within $\tau_F \sim 10^{-8}$ s. For phosphorescence to occur, the excited electron must undergo intersystem crossing to a triplet state, a radiationless process that causes inversion of the electronic spin. Hence, the electron in the triplet excited state and the electron remaining in the ground state have then identical orientation. Vibrational relaxation is then followed by radiative decay to $S_0$. Since this decay involves a quantum-mechanically "forbidden" transition, phosphorescence is less likely to happen than fluorescence, which manifests in decay constants that are higher by several orders of magnitude ($10^{-3} \leq \tau_P \leq 10^0$ s) [109].

### 3.2.2 Characteristics of fluorescence emission

In fluorescence, the absorbed photon (h$\nu_A$) is usually higher in energy than the emitted photon (h$\nu_F$), a phenomenon referred to as Stokes shift (Figure 3.8) [110]. The shape of fluorescence emission spectra is generally not affected by the excitation wavelength, even though there are

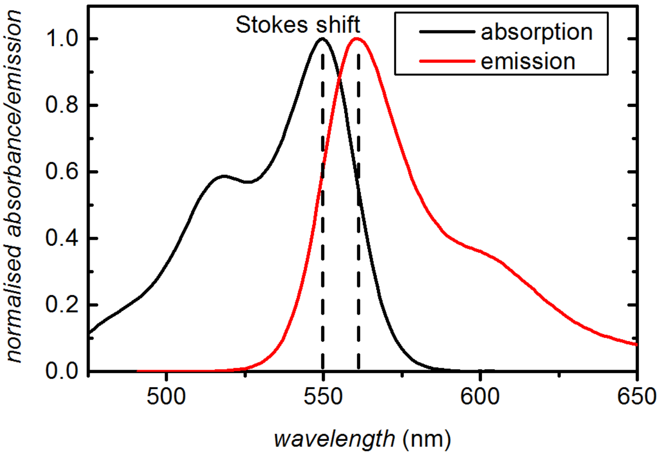


Figure 3.8: Normalised absorption and emission spectrum of sulfonated Cy3 NHS ester in aqueous solution. Spectra recorded in 50 mM MOPS, 100 mM $KNO_3$, pH 6.90. Data recorded by Dr. Danny Kowerko, University of Zurich.

some exceptions [108]. As the same transitions are usually involved in absorption and fluorescence, the corresponding spectra are in many cases well described as mirror images (Figure 3.8) [109]. The *brightness* of a fluorophore is determined by its extinction coefficient ($\varepsilon$) and the quantum yield ($QY$) [111]. $\varepsilon$ is defined by the absorption cross section, which is a measure for the probability that an electron in the ground state absorbs a photon at the excitation wavelength $\lambda_{\text{exc}}$. The quantum yield is the ratio of emitted and absorbed photons:

$$QY = \frac{k_{\text{r}}}{k_{\text{r}} + k_{\text{nr}}} \tag{3.1}$$

As a consequence, the $QY$ depends both on the emissive rate of the fluorophore (rate of relaxation concomitant with the emission of a photon, $k_{\text{r}}$), and its rate of non-radiative decay ($k_{\text{nr}}$) [109]. Experimentally, the $QY$ is determined by measuring the fluorescence of a reference fluorophore of known quantum yield $QY_{\text{ref}}$ using identical experimental parameters, *i.e.* fluorophore concentration, buffer condition, excitation wavelength, slit width, and photomultiplier voltage.

$$QY = QY_{\text{ref}} \frac{I}{I_{\text{ref}}} \frac{1 - 10^{-OD_{\text{ref}}}}{1 - 10^{OD}} \frac{\eta^2}{\eta_{\text{ref}}^2} \tag{3.2}$$

Here, $I$ denotes the area under the fluorescence emission peak, $OD$ refers to the optical density at the excitation wavelength (also: *absorbance*), and $\eta$ is the refractive index of the medium [112]. The average time to decay from the excited state $S_1$ to the ground state is denoted fluorescence lifetime ($\tau_{\text{F}}$) and like the $QY$, it depends on $k_{\text{r}}$ and $k_{\text{nr}}$ [113].

$$\tau_F = \frac{1}{k_{\text{r}} + k_{\text{nr}}} \tag{3.3}$$

Hence, both the $QY$ and $\tau_{\text{F}}$ will respond to factors that alter either of the decay constants,

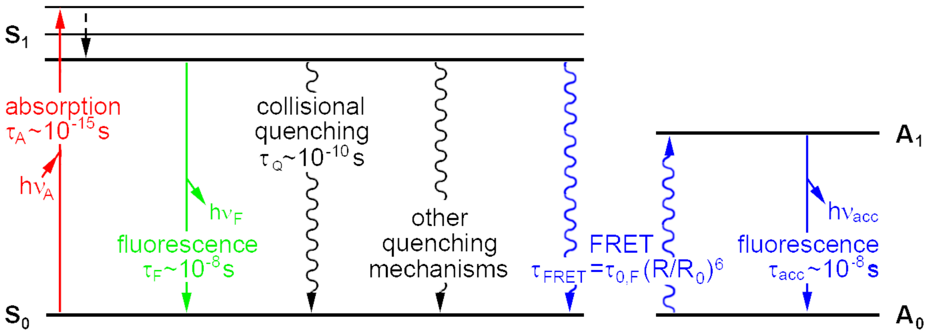


Figure 3.9: Jablonski diagram depicting fluorescence quenching through different mechanisms. Collisional quenching (also: *dynamic quenching*) occurs by diffusional contact of the excited fluorophore with an appropriate quencher. Non-radiative decay via FRET involves energy transfer to a suitable acceptor, which is usually (but not necessarily) a fluorophore. Other quenching mechanisms lead to non-radiative relaxation to the ground state $S_0$. Scheme adapted from [109].

such as quenching or resonance energy transfer (see also: Subsections 3.2.3 and 3.2.3) [113]. Different mechanisms of non-radiative decay are described in the following sections.

### 3.2.3 Fluorescence quenching

Fluorescence quenching refers to any process that decreases the emission intensity of a fluorophore and can be caused by a variety of molecular interactions [109]. Collisional quenching (Figure 3.9) consists of deactivation of the excited fluorophore upon the diffusive encounter with another molecule and generally adheres to the Stern-Vollmer equation [114, 115]:

$$\frac{I_0}{I} = 1 + Kc(\mathrm{Q}) = 1 + k_{\mathrm{Q}}\tau_0 c(\mathrm{Q}) \tag{3.4}$$

where $I_0$ and I denote the absolute fluorescence intensities in the absence and presence of the quencher, $K$ is the Stern-Vollmer constant, $c$(Q) is the quencher concentration, $\tau_0$ is the fluorescence lifetime in the absence of the quencher [109], and $k_\mathrm{q}$ is the bimolecular quenching constant. Here, large values of $k_\mathrm{q}$ should be interpreted as efficient quenching by Q and good accessibility of the fluorophore to Q. A number of molecules are known to act as collisional quenchers, including molecular oxygen, amines, and transition metal ions [116, 117, 118]. Other molecular interactions leading to decreased fluorescence are excited-state reactions, molecular rearrangements, ground-state complexation (static quenching) [119], interaction with a surface [120], and energy transfer (see also: next section) [109].

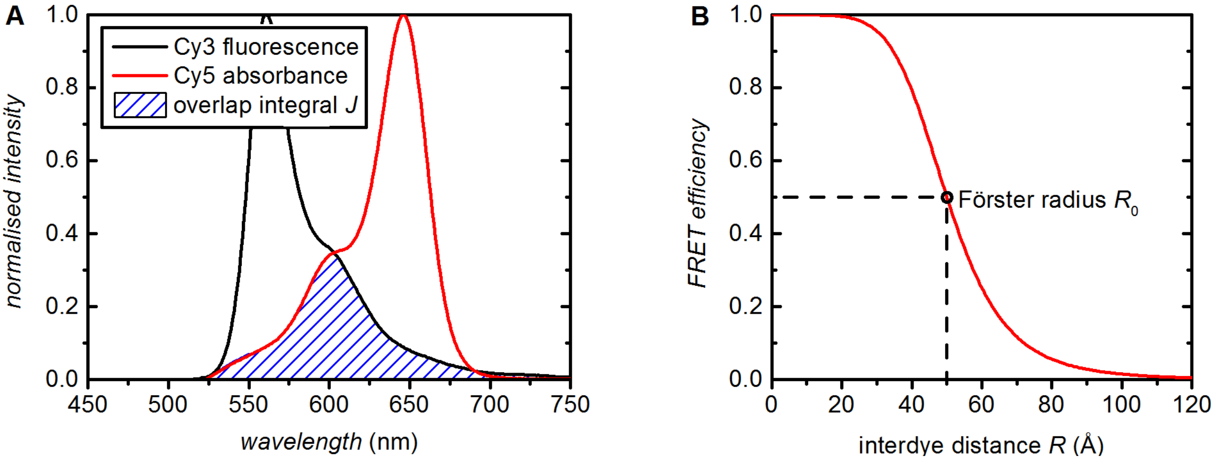


Figure 3.10: (A) Normalised emission and absorption spectra of sulfonated Cy3 and sulfonated Cy5 in 50 mM MOPS, 100 mM $KNO_3$, pH 6.90. Data recorded by Dr. Danny Kowerko, University of Zurich. (B) Dependence of the FRET efficiency on interdye distance. $R_0$ denotes the Förster radius at which $E_{\mathrm{FRET}} = 0.5$ (Equation 3.6).

### 3.2.4 Förster resonance energy transfer

Förster resonance energy transfer (FRET) is defined as energy transfer via a long-range dipole-dipole interaction that occurs between a donor fluorophore and an acceptor (Figure 3.9) [109]. FRET results in a decrease in both donor emission intensity and lifetime, and the appearance of acceptor fluorescence (if the latter is fluorescent) [121]. The efficiency of the process is given by:

$$E_{FRET} = \frac{1}{1+(R/R_0)^6} \tag{3.5}$$

where $R$ is the distance between the donor and the acceptor and $R_0$ is the so-called Förster distance at which 50% of the energy is transferred (also: emphFörster radius) [122]. $R_0$ is formally defined as:

$$R_0 = 9.78\ 10^3(\kappa^2\eta^{-4}QY_{\text{don}}J(\lambda)) \quad (in\ Angstrom) \tag{3.6}$$

where $\kappa^2$ denotes the relative orientation of the transition dipoles of the donor and the acceptor, $\eta$ is the refractive index of the medium, $QY_{\text{don}}$ is the quantum yield of the donor, and $J(\lambda)$ is the overlap integral between the donor emission and the acceptor absorption spectra (Figure 3.10A) [123].

Owing to a steep distance dependence on a biologically relevant scale, FRET is regularly referred to as a *spectroscopic ruler* and has been used to study biomolecular structure and dynamics (Figure 3.10B) [28, 124, 125, 126, 127]. However, for precise determination of interdye distances, $R_0$ has to be calculated. While a considerable body of $\eta$ values has been precedented in the literature and the determination of $QY_{\text{don}}$ and J from absorption and emission spectra is straightforward (Figure 3.10A), the value of $\kappa^2$ is often associated with uncertainty [128]. Both donor and acceptor fluorophore are typically assumed to rotate freely, which corresponds to a $\kappa^2$ value of 2/3. In practice, however, rotation of fluorophores has frequently been proven hindered upon attachment to biomolecules, and $\kappa^2$ may then adopt values between 0 and 4 [129, 130]. Under such circumstances, changes in relative dye orientation and changes in their distance cannot be discerned, an issue sometimes referred to as the $\kappa^2$ problem [131]. As a consequence, control experiments addressing the rotational behavior of the fluorophore must be performed to rule out experimental artifacts (*fluorescence anisotropy*, see also: Section 3.2.5). It should also be noted that application of FRET in biophysical research do often not require knowledge about absolute distances, as it is also the case here. In practice, so-called apparent FRET efficiencies are often sufficient, circumventing many further issues, a detailled description of which would go beyond the scope of this work.

### 3.2.5 Fluorescence anisotropy

When linearly polarised light is used for fluorophore excitation, only a subset of molecules with their dipole moments oriented along the electric vector of the excitation will be excited (photoselection, Figure 3.11) [121]. The emitted light is also polarised along a fixed axis in the fluorophore, which corresponds to the orientation of the emission dipole [109]. The relative angle between the excitation polarisation and the emission polarisation is used to calculate fluorescence polarisation ($P$) and anisotropy ($r$):

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \tag{3.7}$$

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{3.8}$$

with $I_{\parallel}$ the emission intensity along the excitation polarisation and $I_{\perp}$ the emission intensity perpendicular to the excitation polarisation [121]. In practice, anisotropy is often preferred over polarisation, as the former is normalised by the total emission intensity and $P$ can easily be rewritten as $r = 2P/(3\text{-}P)$ [132, 133]. The value of $r$ depends upon the ratio of the rates of angular displacement of the dye, its rotational diffusion and the excited state lifetime and adopts 0 in the case of total depolarisation (Figure 3.11 [134]. When fluorophores are dissolved in a fluid solution, they typically rotate extensively in $\sim 10^{-10}$ s, which is considerably shorter than the excited state lifetime ($\tau_F \sim 10^{-9}$ s, see section 3.2.1) [109]. However, conjugation of the fluorophore to a macromolecule, surface immobilisation, as well as viscosity are known to reduce the rate of rotational diffusion and may lead to an increase in $P$ and $r$ [135, 136, 137]. Altered rotational diffusion has to be identified as such in order to avoid misinterpretations of the experimental data. In practice, one distiguishes steady-state (also: *static*) anisotropy measurements and time-resolved (also: *dynamic*) anisotropy experiments, the latter allowing to directly monitor depolarisation [138].

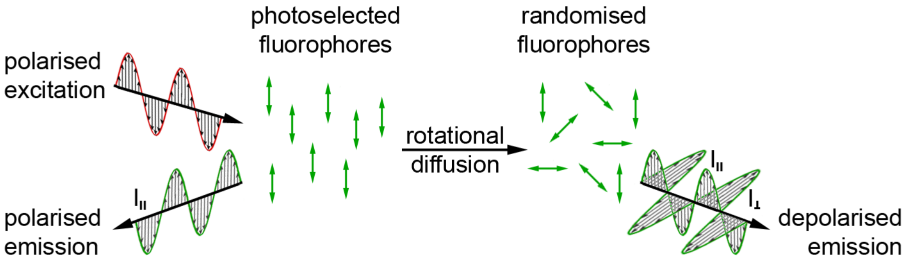


Figure 3.11: The principle of photoselection by linearly polarised light and the of emission polarisation on rotational diffusion. Scheme adapted from [109].

### 3.2.6 Commonly used fluorophores and FRET pairs

Ideal fluorophores for adressing biological questions must display a high extinction coefficient ($\varepsilon > 50'000\ \mathrm{M}^{-1}\ \mathrm{cm}^{-1}$) and quantum yield ($QY > 0.1$) [111]. They are further required to be photostable and undesirable photophysical, chemical and aggregation effects should be largely absent [141]. As the majority of biological assays is carried out in aqueous solutions, they must also be water-soluble [141]. Finally, fluorophores should be small to induce minimal perturbation of the system under study and they should have functionalities for straightforwards bioconjugation [3, 139]. Additionally, a good FRET pair has pronounced spectral separation between donor and acceptor emissions for reduced crosstalk, *i.e.* detection of donor emission in the acceptor channel and direct excitation of the acceptor [141]. Similar quantum yields and detection efficiencies (this may be problematic for dyes emitting in the far-red region) are also desirable in order to unambiguously identify anticorrelation, unless the FRET acceptor consists of a dark quencher [142].

Commonly used fluorophores are shown in Figure 3.12 and are generally structurally related, *i.e.* they typically contain conjugated double bonds with delocalised $\pi$ electrons. In general, increasing conjugation generally shifts the wavelength of absorbed light towards higher values, which allows to tune the spectral properties to a certain extent. Pyrene, naphthalene and coumarin derivatives tend to emit in the UV region, while fluorescein, rhodamine, and cyanine dyes typically emit in the vis/near-IR spectrum [139]. However, emission spectra can be

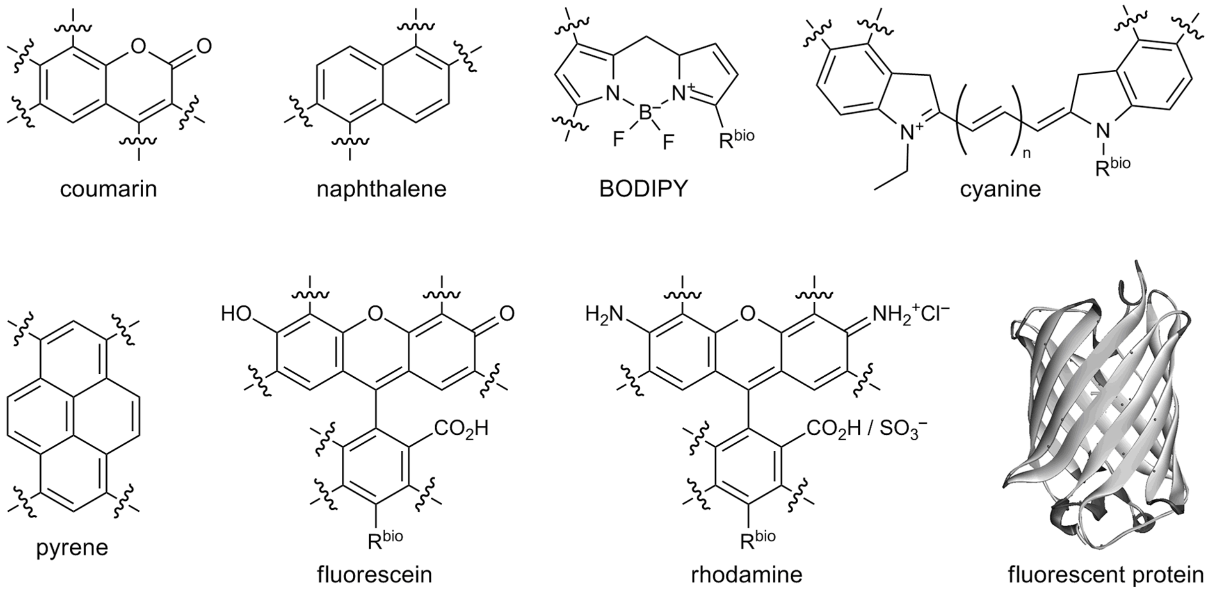


Figure 3.12: Structures of common UV/vis fluorophores. Substituents at the $\sim$ position include $CH_3$, OH, $CO_2^-$, $OCH_3$, $NO_2$ and $SO_3^-$ are incorporated to spectrally shift absorption/emission spectra, improve water-solubility, and alter the photophysical bevavior. $R^{bio}$ indicates the typical position of the linker for bioconjugation. Scheme adapted from [139]. The crystal structure depicts the *Aequorea victoria* green fluorescent protein and was taken from [140].

shifted depending on the incorporated functionalities. Numerous derivatives of the wild-type green fluorescent protein (GFP) from *Aequorea victoria* are nowadays available, which span emissions from 470 nm to 650 nm [143, 144]. Quantum dots, *i.e.* nanocrystals made from semiconductor materials, may also act as fluorophores [145]. It is important to note that each fluorophore has its own personality regarding spectroscopic and chemical properties [146]. For example, fluorescein dyes have high quantum yields and solubility and can be readily conjugated to biomolecules [139]. In turn, they are also known to display high photobleaching rates and their quantum yields are pH-dependent [147]. Cyanine dyes are very popular choices for studying nucleic acid conformations, because their photostability can be improved through removal of molecular oxygen and by the addition of a suitable triplet-state quencher such as the vitamin E analogue Trolox [146, 148, 149, 150, 151]. This work uses the cyanine dyes Cy3 and Cy5. Please refer to the Appendix (Section 12.2) for chemical structures including the linker and a survey of their spectral properties.

### 3.2.7 Importance of fluorescence in biological research

Since fluorescence-based assays are often non-invasive, they are nowadays crucial in biological research. In particular, fluorescence has widely replaced radioactivity, because it is safer, more convenient, and the straightforward implementation of multichannel measurement. Examples are:

- **Agarose gel electrophoresis:** Intercalating agents are routinely used as tags to detect nucleic acid bands via fluorescence. Commonly used compounds are ethidium bromide and the cyanine dyes SYBR®Green I and SYBR®Gold [152].

- **Sanger sequencing:** Whereas Sanger sequencing originally employed radioactively labelled dideoxynucleoside 5'-triphosphates (ddNTPs), they have nowadays been widely replaced by fluororescently labelled ddNTPs [153, 154].

- **Microarrays:** DNA microarrays use fluorophore-labelled nucleic acid samples (typically Cy3 and Cy5), which hybridises to the probes in a sequence-specific manner, allowing for gene profiling and the detection of single nucleotide polymorphisms.

- **Quantitative polymerase chain reaction:** A quantitative polymerase chain reaction (qPCR) is an extension of a regular PCR, in which non-specifically intercalating dyes, for example SYBR®Green I, are utilised to detect and quantify the PCR product [155].

- **Immunofluorescence techniques:** Immunofluorescence uses fluorophore-tagged antibodies to visualise target molecules in a sample. A number of organic fluorophores can be coupled to the antibody, including aminomethylcoumarin, Cy2, fluorescein, Cy3, tetramethyl rhodamine, rhodamine red-X, Texas Red, and Cy5 [156]. Applications include western blotting, fluorescence-activated cell sorting (FACS), and fluorescence *in situ* hybridisation (FISH) [157, 158, 159].

- **Fluorescent proteins as reporters of expression:** Due to their low phototoxicity and the possibility to fuse it to a target protein, fluorescent proteins have found wide application in molecular and cell biology as a reporter system [160].

## 3.3 Measuring fluorescence, one molecule at a time

Single-molecule fluorescence techniques are dedicated to measuring the fluorescence emission of single dyes. Such experiments in principle provide the same information as a standard ensemble measurement, as a sufficiently high number of observations from a single molecule is equivalent to a snapshot of the whole population (ergodicity hypothesis) [161]. However, "looking at just one thing" has a number of advantages [162] (list adapted from [161]):

- **Heterogeneity and disorder:** Single molecule methods unveil heterogeneity and disorder within a sample, *i.e.* individual molecules do not behave identically during the window of observation. Such cross-sample variability has frequently been reported in single-molecule studies, albeit it does (usually) not contradict the principle of ergodicity [163, 164]. Examples of cross-sample variability, possible origins, and a method for its quantification are described in Chapters 5, 6, and 8.
- **Reaction intermediates:** Single-molecule approaches in principle reveal the whole pathway of a reaction, including transient intermediates and parallel reactions [165].
- **Quantification of kinetics:** The rate constants associated with complex biological processes can be determined, including nucleic acids and protein folding reactions and the displacement of molecular motors along their substrates [166]. Further examples are discussed in detail in Chapter 4.
- **Precise localisation and counting:** So-called superresolution imaging techniques achieve resolutions that are not limited by the diffraction of light, hence allowing to localise fluorophores with nanometre accuracy [167].
- **Low concentrations:** Performing experiments with a very low number of molecules (1-$10^4$) permits to work at near-physiological concentrations ($10^{-12}$ - $10^{-9}$ mol/L) and reduces costs [168].

### 3.3.1 Microscope designs for detection of single-molecule fluorescence

Single-molecule spectroscopy relies on the detection of single photons at energies in the range of just $10^{-19}$ J. Therefore, instruments generally confine the excitation volume and/or reject background noise that can originate from autofluorescence, and elastic Rayleigh and inelastic Raman scattering [161, 169]. Detection of single fluorophore emission generally use either wide-field or confocal geometry, which differ in the way they excite the sample. In wide-field microscopy, the entire specimen is illuminated at once, while confocal microscopy uses point illumination. This section provides an overview of common microscope setups used to perform single-molecule fluorescence experiments.

### Wide-field microscopy: total internal reflection fluorescence microscopy

Total internal reflection fluorescence microscopy (TIRFM) generates an evanescent field, reducing the excitation volume to a thin sheet at the quartz-solution interface [170]. This incident laser beam must reach the sample in a highly inclined manner, such that the angle $\theta > \theta_{\text{crit}}$, the so-called critical angle (Figure 3.13A) [170]. According to Snell's Law, $\theta_{\text{crit}}$ is defined as

$$\theta_{\text{crit}} = \frac{1}{sin(\frac{\eta_{\text{buf}}}{\eta_{\text{q}}})} \tag{3.9}$$

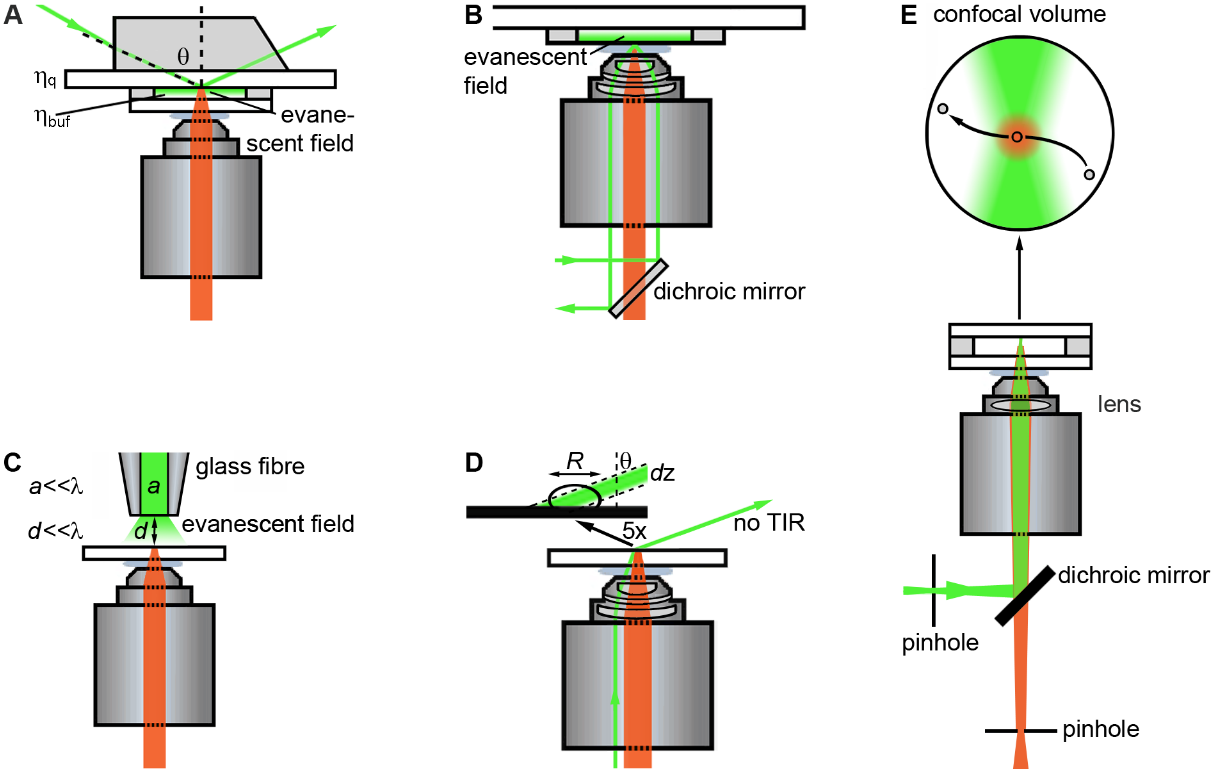


Figure 3.13: Non-exhaustive overview on microscope systems suitable for single fluorophore detection. (A) Prism-based total internal reflection fluorescence microscopy. $\theta$ denotes the angle of the incident excitation light, $\eta_{\text{q}}$ and $\eta_{\text{buf}}$ are the refractive indices of quartz and the buffer. (B) Objective-based total internal reflection microscopy. (C) In near-field scanning optical microscopy, both the diameter of the aperture $a$ and the distance between the capillary and the sample $d$ are smaller than the wavelength of the light used for excitation $\lambda$. (D) Highly inclined and laminated optical sheet microscopy. Green represents the excitation beam, orange fluorescence emission. (E) Laser confocal microscopy. The cartoon depicts the diffusion of a fluorophore and its emission within the confocal volume. Schemes adapted from [161, 170, 171, 172].

where $\eta_{\mathrm{buf}}$ and $\eta_{\mathrm{q}}$ denote the refractive indices of the solution and quartz [173]. For example, $\theta_{\mathrm{crit}}$ at a quartz-water interface approximates 69.93°. Pratical implementations of TIRFM employ either a quartz prism (Figure 3.13A) or the objective (Figure 3.13B) of the microscope to create an evanescent field and illuminate surface-tethered or surface-constrained fluorophores [161]. Single-molecule FRET (smFRET) experiments on surface-immobilised nucleic acids, including all smFRET experiments presented in this work, are typically conducted using TIR excitation. Please refer to Section 12.1 in the Appendix for a detailled description of the microscope used.

### Widefield microscopy: near-field scanning optical microscopy

Near-field scanning optical microscopy (NSOM) uses a lens-free optical fibre for excitation [171]. Importantly, it is equipped with an aperture $a$ of sub excitation wavelength diameter ($a << \lambda$), which generates an evanescent field that is then operated at distances below the excitation wavelength ($d << \lambda$, Figure 3.13C). The image is recorded either through the same fibre or using a high numerical aperture objective [161]. NSOM was used in pioneering studies demonstrating that it is possible to detect fluorescence of single dyes, as well as FRET between single fluorophores attached to a nucleic acid support [174, 175, 176]. However, it is nowadays less frequently employed, because (i) multiple molecules can only be imaged in scanning mode and (ii) scanning drastically reduces the temporal resolution and may easily cause the tip to break [177].

### Widefield microscopy: highly inclined and laminated optical sheet microscopy

In highly inclined and laminated optical sheet (HILO) microscopy, the incident laser light is highly inclined (but: $\theta < \theta_{\mathrm{crit}}$) and refracts into a thin optical sheet [161, 172]. This greatly minimises the illumination volume to roughly $R/\tan\theta$ in the $z$ direction [172] (Figure 3.13D). HILO microscopy has been sucessfully implemented to image single fluorophores on the surface of cells, as well as for single-molecule tracking inside living tissue [172, 178]. In turn, applications in single-molecule FRET have not yet been reported, which might be due to the fact that TIRF microscopy achieves higher signal-to-noise ratios than a HILO geometry [179].

### Confocal microscopy

In laser confocal microscopy (LCM), pinholes are placed in both the excitation and emission path (Figure 3.13E) [161]. Thus, the excitation volume is confined to a diffraction-limited Gaussian beam waist and out-of-focus fluorescence is rejected, which greatly suppresses background noise [161, 180]. LCM has been applied to detect single fluorophore emssion and fluorescence polarisation [175, 181], in fluorescence lifetime imaging microscopy (FLIM) [182], and in smFRET experiments [183, 184]. In particular, LCM is commonly used to detect fluorescent bursts from fluorophore-labelled biological macromolecules freely diffusing in solution, followed by calculating the FRET efficiency in order to characterise the distribution of conformations and structural dynamics [185].

#### Light sources and detectors

Usually, lasers are the preferred light sources for single-molecule fluorescence applications, because they provide a defined excitation wavelength, as well as high and stable intensities [161]. Depending on the experimental requirements, they can be operated either continuously or pulsed (see also: Section 12.1.2 in the Appendix). Common detectors for wide-field microscopes are electron-multiplying charge-coupled device (EM-CCD) cameras that achieve high quantum efficiencies in the visible range, low readout noise, fast vertical shift speed, and relatively high temporal resolution ($\sim$ 10 ms) [141]. In turn, single photon-counting avalanche photodiodes (APDs) are typically used in conjunction wirh a confocal geometry and they combine high readout rates and high quantum efficiencies [141]. Photomultiplier tubes (PMTs) have larger detection areas, but display lower quantum yields and have to be operated at higher voltage, which leads to an increased number of dark counts [161]. Hybrid photomultiplier detectors combine large detection areas, high readout rates ($\sim$ 50 ps), high detection efficiencies, and decreased photophysical artifacts (afterpulsing) and are nowadays widely implemented in commercially available microscopes [186].

### 3.3.2 EM-CCD data processing and analysis

The majority of smFRET studies on RNA folding and catalysis have been conducted using an EM-CCD for photon detection, resulting in time-binned intensities. Raw movies usually resemble the averaged example given in Figure 3.14A and must be processed further. This section gives an overview on the stepwise analysis of such raw movies in order to extract single-fluorophore *emission* and *FRET* trajectories, which are subsequently used to characterise the conformational equilibrium and the rate constants associated with a conformational change of a biological macromolecule.

#### Automatic localisation of fluorophores in a crowded field

As fluorophores are much smaller than the spatial resolution of the microscope ($\sim 10^{-8}$ m$^2$ as estimated by Mélodie Hadzic, University of Zurich), single fluorophore emission appears as a so-called point-spread function (PSF, Figure 3.14B). In smFRET experiments, the detection field of the CCD chip is typically severely crowded by $>100$ single molecules and it would be time-consuming to localise them manually (*crowded field analysis*, [187]). The fact that PSFs are generally well approximated by two-dimensional Gaussian fit functions can be used to automate their localisation as depicted in Figure 3.14B [124]. Other strategies also exist: for example, the averaged image can be transformed into a hough matrix, followed by peak detection using arbitrarily defined parameters such as minimum average intensity and minimum inter-spot separation [188].

#### Co-localisation of FRET pairs

The coordinates identified have to be co-localised, *i.e.* two PSFs that stem from the emission of fluorophores attached to the same biomolecule have to be identified as such. Usually, an averaged image from surface-immobilised fluorescent microspheres serves as a calibration tool to quantify the offset and the distortion between the two detection channels by manually selecting a number of corresponding spot pairs [125]. Subsequently, coordinates in the donor channel are mapped onto the acceptor channel and *vice versa* using for example a non-linear polynomial or a local weighted mean transformation [170]. Finally, spots that are not clearly separated are removed by defining an arbitrary maximum overlap criterion.

#### Creation of raw trajectories

Raw trajectories are created by integrating the intensity within an area of $n$ x $n$ pixels around the colocalised coordinates (Figure 3.14C, inset). The size of the integration area is defined arbitrarily and depends for example on the spatial resolution and the presence of severely blended PSFs, albeit the latter problem can be overcome by defining an arbitrary maximum overlap criterion (*vide supra*). Raw time traces consist of donor and acceptor emission intensities as a function of time $I_{\mathrm{D}}$ and $I_{\mathrm{A}}$ (Figure 3.14D).

#### Correcting for background noise and crosstalk

Background signal in raw time traces is caused by various processes, including autofluorescence, focal drift, electron multiplication, and read-out in CCD cameras (a comprehensive list is given in Section 8.1) [189]. As the background signal may fluctuate in a local and time-dependent manner, background correction is not trivial. One approach is to divide the image into a number of sub-images, followed by determination of local background in real time using the least intense pixels in every frame (aperture photometry approach) [190].

Fluorophore emission should also be corrected for crosstalk between the two detection channels, *i.e.* leakage of donor emission into the acceptor channel (bleedthrough) and direct excitation of the acceptor at the wavelength used for donor excitation. Crosstalk has to be quantified experimentally and adjusted of raw time traces yields corrected photon counts over time $I_{\mathrm{D,corr}}$ and $I_{\mathrm{A,corr}}$ [191]. For further accuracy and subsequent estimation of distances, corrected intensity time traces can be adjusted for differences in fluorophore quantum yield, wavelength-dependent detection efficiency of the CCD camera, as well as further instrument-dependent factors ($\gamma$ factor) [191]. However, $\gamma$ corrections were not performed in the context of this work.

### Further analysis

Apparent *FRET* efficiency over time is typically calculated from background- and crosstalk-corrected donor and acceptor intensities (Figure 3.14E) [123]:

$$FRET = \frac{I_{\text{A, corr}}}{I_{\text{D,corr}} + I_{\text{A,corr}}} \tag{3.10}$$

These smFRET time traces in principle allow to determine the number of conformational states in the system, their relative occurrence (thermodynamic analysis), as well as the rates at which they interconvert (kinetic analysis). However, distance estimations using apparent *FRET* values should be interpreted with extreme care. Depending on the signal-to-noise ratio and the separation of the states, further processing steps may be performed, for example denoising or hidden-Markov-based analysis, followed by characterisation of the thermodynamic equilibrium and quantification of rate constants associated with conformational changes [166, 192]. Different strategies for analysing noisy time traces and commonly used approaches in thermodynamic and kinetic analysis are reviewed in detail in Section 8.1.

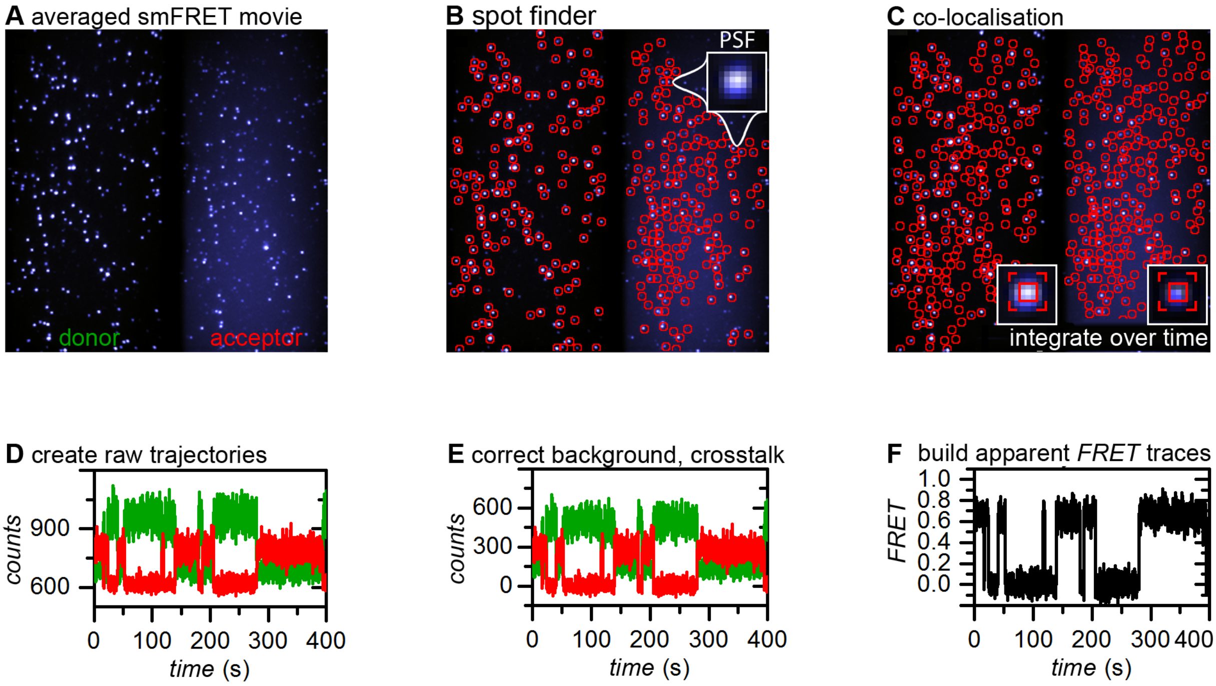


Figure 3.14: Creation of smFRET time traces from raw data. (A) Side-by-side image of spectrally separated single fluorophore emission. (B) Automatic spot detection using the Matlab built-in function *houghpeaks*. The inset illustrates that single fluorophore emission appears as a point spread function that can be approximated by a two-dimensional Gaussian fit. (C) Mapping the coordinates identified in the donor channel onto the acceptor channel and *vice versa* (*co-localisation*) is used to find corresponding spot pairs that are subsequently integrated in the original (unaveraged) movie. (D) Raw trajectory of donor and acceptor emission over time. (E) Subtraction of background noise and correction for crosstalk yields a corrected time trace that can be used to calculate (F) FRET over time according to Equation 3.10.

### 3.3.3 Application of smFRET in studying RNA folding and function

FRET between single, surface-immobilised fluorophores was first demonstrated by Ha *et al.* in 1996, followed by detection using freely diffusing molecules by Deniz and co-workers in 1999 [175, 183]. Numerous applications of smFRET in studying functional RNAs have since been published, capturing RNA folding and/or catalysis in unprecedented detail. Key studies are summarised in this section.

One of the first smFRET studies on RNA involved surface-immobilised **three-way junctions derived from ribosomal RNA** (rRNA, Figure 3.15) [176]. The conformational equilibrium and dynamics were characterised in response to binding of the ribosomal protein S15 and changes in $Mg^{2+}$ concentration [176]. While the results closely resembled the averaged behavior of unlabelled RNA in solution, thus further validating the surface-immobilisation approach, they also revealed $Mg^{2+}$-induced conformational heterogeneity at 100-1000 $\mu$M $Mg^{2+}$.

The hairpin ribozyme plays an important role in the rolling cycle replication of the tobacco ringspot virus and has been extensively studied by smFRET. In its natural form, it is comprised of a four-way junction and two internal loops situated on two adjacent helical arms, the latter of which are required to interact (dock) for catalysis to occur (Figure 3.15) [194]. Using a minimal catalytically active construct devoid of the junction (**minimal hairpin ribozyme**), Zhuang *et al.* showed that docking and undocking of individual ribozymes can be followed in real time using smFRET (Figure 3.15) [195]. Analysing these data, they demonstrated that interdomain docking was governed by single-exponential kinetics, while undocking was highly heterogeneous with four rate constants spanning three orders of magnitude [195]. Heterogeneity was also associated with catalytic activity, as a minimum of two rate constants were required to satisfactorily describe the experimental data [195]. Follow-up studies generally confirmed the unconventional kinetic behavior, even though the values of the rate constants were not always reproduced and docking was also sometimes described as a multiexponential process. By performing site-specific base mutations and $\phi$ value analysis, Bokinsky *et al.* demonstrated that native tertiary contacts are (almost) non-existent in the activated complex [196]. However, $Na^{+}$ and $Mg^{2+}$ titration experiments in conjunction with computational simulations suggested that structural compaction occurs already at this stage. Rueda and co-workers showed that functional single nucleotide modifications, even though they are distal from the catalytic site and not directly involved in the cleavage reaction, affect both docking, undocking, and cleavage rates [197]. Ditzler *et al.* showed that the minimal hairpin ribozyme migrates as two separate bands in native gel electrophoresis, which differ with regard to the undocking kinetics [198]. Molecular heterogeneity was surprisingly resistant against denaturation and did not stem from chemical differences detectable by mass spectrometry (isomers are generally not distinguishable by mass spectrometry).

Tan and co-workers studied a four-way junction derived from the **natural hairpin ribozyme** (Figure 3.15) [199]. They found that in contrast to the minimal construct, both docking and undocking are multiexponential processes, a phenomenon that was reproduced upon surface passivation with BSA, PEG, and upon encapsulation of RNA within surface-tethered vesicles [199, 200]. Tan *et al.* also reported that compared to the two-way junction, docking and undocking occurs faster by three orders of magnitude, an observation that could be explained by a previously unprecedented obligate folding intermediate in the presence of the junction

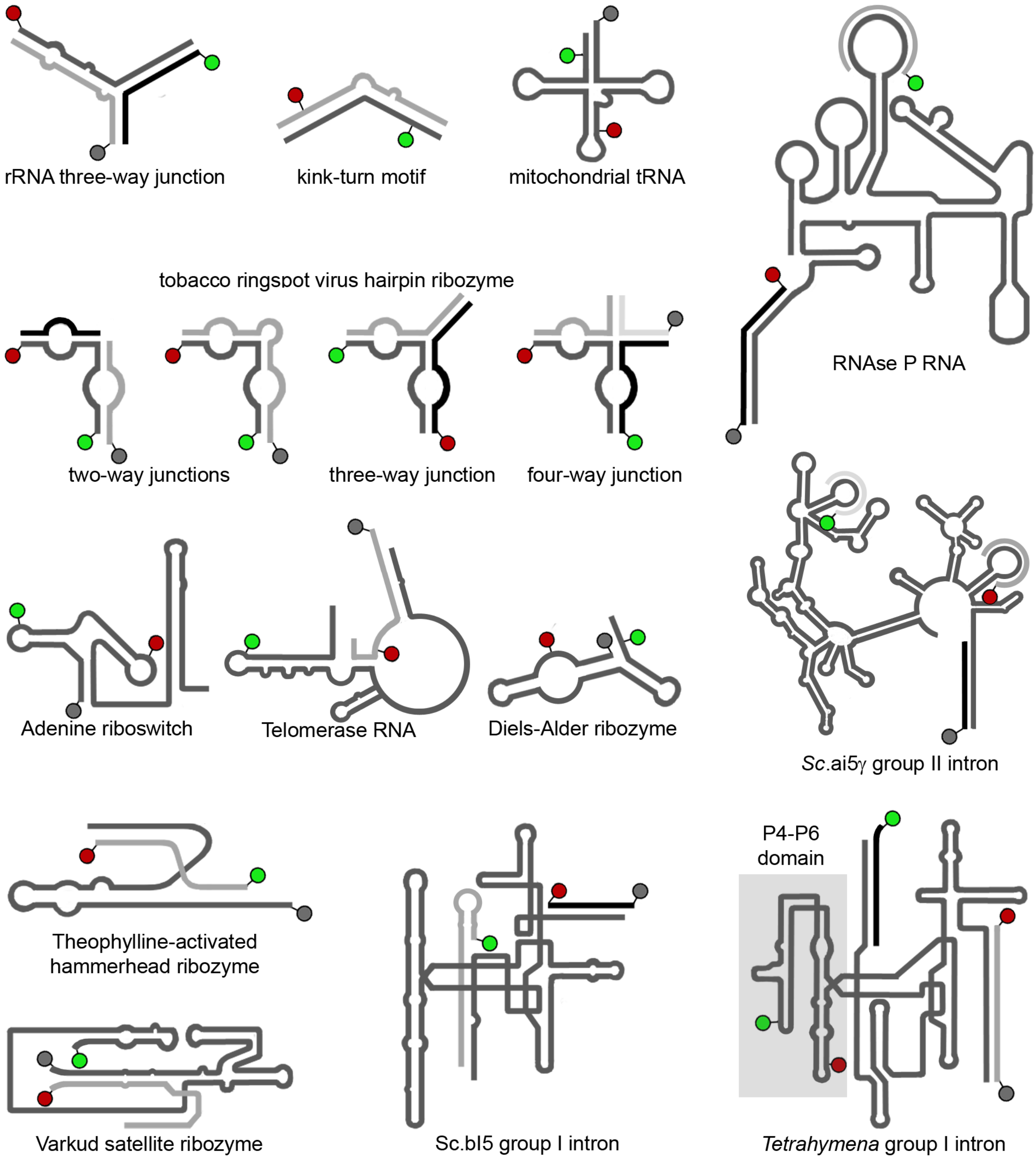


Figure 3.15: Non-exhaustive scheme of RNA molecules that have been studied using smFRET. Green and red spheres indicate the position of the FRET donor and acceptor fluorophores, respectively. When applicable, grey spheres represent biotin moieties for surface immobilisation. Figure adapted from [193].

[199]. The presence of a four-way-junction-specific folding intermediate was confirmed by Pljevaljcic *et al.*, albeit it displayed considerably higher *FRET* efficiency ($\sim 0.85$ vs $\sim 0.50$) and formed at two to four orders of magnitude higher rates [194]. The conformational equilibrium of the isolated four-way junction with perfectly base-paired helical stems (no loops) was studied by Hohng *et al.* who showed that conformational fluctuations of the undocked ribozyme are due movement of the junction [201]. Nahas and co-workers further characterised conformational dynamics and chemistry of the fully functional natural ribozyme [202]. They found out that in the absence of single base modifications aimed at inhibiting cleavage activity, docking and undocking occur at rates lower by about one order of magnitude [199, 202]. They also reported discrepancies between the internal cleavage rates of the natural and the minimal ribozyme [195, 197, 202], which were subsequently resolved by Liu *et al.* who performed so-called kinetic fingerprinting experiments to show that like for the four-way junction, utilisation of a non-cleavable substrate artefactually increases the rate of undocking by a factor of 20 [203]. Taken together, smFRET studies of the hairpin ribozyme have not only helped validating experimental and analytical approaches that are nowadays widely employed, they have also provided a picture RNA folding and catalysis in unprecedented detail.

To date, one of the largest functional RNA studied by smFRET is the D135-L14 ribozyme derived from the **yeast group II intron ai5$\gamma$**. Steiner *et al.* reported a previously unprecedented obligate intermediate along the folding pathway leading to a new folding paradigm [204]. Steiner and co-workers also performed the first investigation of a divalent metal ion other than $Mg^{2+}$ on RNA folding, thereby demonstrating that inhibition of splicing in the presence of $Ca^{2+}$ previously reported in bulk experiments concurs with a division of single molecules into two subpopulations that display distinct *FRET* traces [78, 87]. These findings were interpreted as differences regarding the global RNA fold mediated in the presence of $Ca^{2+}$ [87]. Additionally, Karunatilaka *et al.* showed that the DEAD-box helicase Mss116 lowers the energy barrier between different conformational states and that it promotes RNA folding via an ATP-dependent and an ATP-independent pathway [205].

A multitude of other RNA molecules have been studied by smFRET, a detailled description of which would be beyond the scope of this work. Further examples of small ($<$100 nt) naturally occurring RNA molecules that have been studied by smFRET are **telomerase RNA** [206], **Varkud satellite RNA** [207], the **U4 snRNA kink-turn motif** [208], **U2-U6 snRNA** [209], **bacterial riboswitches** responding to adenine, lysine and cyclic di-GMP [211, 212, 213], the **DsrA-rpoS bulge loop** [214, 215], **mitochondrial tRNA** [216, 217], and **transfer messenger RNA** [218]. The ***Tetrahymena* group I intron** [219, 220, 221, 224] and its P4-P6 domain [222, 223, 226, 225], the phylogenetically close **mitochondrial group I intron bi5** [227], as well as **RNase P RNA** [228, 229, 230] are examples of large ($>$100 nt) naturally occurring ribozymes that have been studied using smFRET. Additionally, *in vitro* selected aptamers and artificial ribozymes have also been studied, such as the **GTP-binding aptamer** [231], the **theophylline-activated hammerhead ribozyme** [232], and the **Diels-Alderase ribozyme** [233].

Table 3.4: Examples of (cation-mediated) RNA folding and catalysis as studied by single-molecule FRET. List adapted from [193].

| Scientific model | Biological context | Cation dependence | References |
|---|---|---|---|
| Three-way junction from *E. coli* rRNA | translation | $Mg^{2+}$ | [176] |
| *S. cerevisiae* ai.5$\gamma$ mitochondrial group II intron | self-splicing | $K^+$, $Mg^{2+}$, $Ca^{2+}$ | [87, 204, 205] |
| Tobacco ringspot virus hairpin ribozyme | rolling cycle virus | $Na^+$, $Mg^{2+}$ replication | [194, 195, 196, 197, 198]<br>[199, 200, 202, 201, 203] |
| *Tetrahymena* telomerase RNA | DNA maintenance | - | [206] |
| *Neurospora* VS RNA | self-cleavage | - | [207] |
| *Human* U4 snRNA kink turn motif | spliceosomal splicing | $Na^+$, $Mg^{2+}$ | [208] |
| *S. cerevisiae* U2-U6 snRNA | spliceosomal splicing | $Mg^{2+}$ | [209, 210] |
| Adenine riboswitch | translation | - | [211] |
| Lysine riboswitch | translation | - | [212] |
| c-di-GMP riboswitch | translation | $Mg^{2+}$ | [213] |
| *E. coli* DsrA-rpoS bulge loop | translation | - | [214, 215] |
| *Human* mitochondrial $tRNA^{Lys}$ | tRNA maturation | $Mg^{2+}$ | [216, 217] |
| Transfer messenger RNA | ribosome rescue | $Mg^{2+}$ | [218] |
| *Tetrahymena* group I intron | self-splicing | $Na^+$, $Mg^{2+}$, $Ba^{2+}$ | [219, 220, 221, 222]<br>[223, 224, 225, 226] |
| *S. cerevisiae* bl5 mitochondrial group I intron | self-splicing | $Mg^{2+}$ | [227] |
| *B. subtilis* RNAse P RNA | tRNA processing | $Mg^{2+}$ | [228, 229, 230] |
| GTP-binding aptamer | - | - | [231] |
| Theophylline-activated hammerhead ribozyme | - | - | [232] |
| Diels-Alderase ribozyme | - | $Mg^{2+}$ | [233] |

In summary, there is a large number of naturally occurring and artificial RNA constructs fundamentally differing in fold and function that have been studied by smFRET (Figure 3.15 and Table 3.4). The majority of these studies employed a prism-based TIR geometry used in conjunction with a CCD detector, most likely because (i) RNA folding is generally slow and the time resolution of the CCD chip is sufficient, (ii) TIRFM allows to image hundreds of single molecules at a time, and (iii) the implementation of a TIR geometry is less challenging than the implementation of a confocal one [234]. Table 3.4 also reveals most smFRET studies addressing RNA folding and catalysis precedented to date either focus on alkali ions and the earth alkaline ion $Mg^{2+}$ or they do not investigate the role of metal ions at all.

## 3.4 Thesis outline

The goal of this work is to characterise the influence of divalent metal ions on the formation of RNA tertiary structure using single-molecule fluorescence, as well as to develop a framework for applying statistical rigor in smFRET experiments. The RNA sequences used for smFRET and bulk biochemical experiments are derived from the 5' splice site recognition

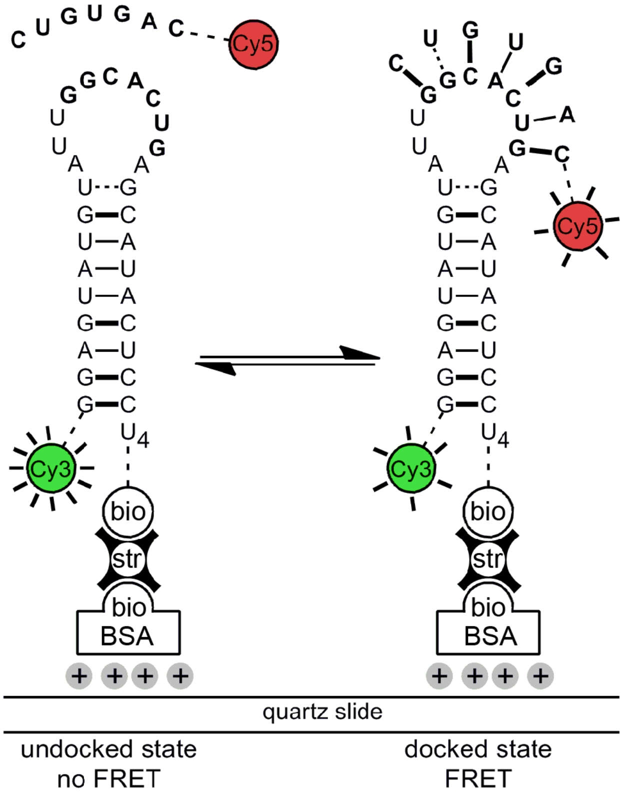


Figure 3.16: Studying d3'EBS1*/IBS1* interaction by smFRET. The d3'EBS1* hairpin is labeled with Cy3 and tethered to the surface of a quartz slide passivated with biotinylated BSA via a biotin-streptavidin linkage. Docking of a Cy5-IBS1* strand is characterised by the appearance of Cy5 fluorescence and a decrease in Cy3 emission due to FRET. Figure and caption taken from [166].

complex within the yeast mitochondrial group II intron *Sc*.ai5$\gamma$ (d3'EBS1* and IBS1*), which are post-synthetically modified with the FRET pair Cy3 and Cy5, as well as a biotin moiety for subsequent immobilisation on a surface (Figure 3.16). *In silico* experiments and statistical analysis are performed using Matlab (version 8.20.701, license 49040, Mathworks, Nattick, MA, USA).

Chapter 4 provides an account of different force- and fluorescence-based single-molecule methods and how they have been utilised to characterise helicase-mediated RNA folding. Chapter 5 addresses the influence of physiological $Mg^{2+}$ and $Ca^{2+}$ on d3'EBS1*/IBS1* interaction using single-molecule fluorescence, followed by a detailled characterisation $M^{2+}$ binding to the RNA by nuclear magnetic resonance. Chapter 6 surveys examples of non-conventional kinetics revealed by single-molecule FRET and summarises their possible origins. Chapter 7 validates smFRET for as a method to systematically study cation-dependent formation of RNA tertiary structure. The metal ion dependent thermodynamics and kinetics associated with interstrand interaction are further shown to be in excellent agreement with complex stabilities along the extended Irving-Williams series, providing a link between the coordination chemistry of the metal ion and RNA structure formation. Finally, Chapter 8 describes the mathematical framwork for using bootstrapping in statistical analysis of smFRET data, followed by assessing the robustness of a home-built bootstrapping software using simulated and experimental data.

# Chapter 4

# Helicase-mediated changes in RNA structure at the single-molecule level

Sebastian L.B. König,[1] Pramodha S. Liyanage,[2] Roland K.O. Sigel[1] and David Rueda[2,3]

1 Institute of Inorganic Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
2 Department of Chemistry, Wayne State University, 5101 Cass Avenue, Detroit, MI 48202
3 Department of Medicine, Imperial College, Du Cane Road, London, SW12 0NN, UK

**ABSTRACT:** RNA helicases are a diverse group of RNA dependent ATPases known to play a large number of biological roles inside the cell, such as RNA unwinding, remodeling, export, as well as degradation. Understanding how helicases mediate changes in RNA structure is therefore of fundamental interest. The advent of single-molecule spectroscopic techniques has unveiled with unprecedented detail the interplay of RNA helicases with their substrates. In this review, we describe the characterisation of helicase-RNA interactions by single-molecule approaches. State-of-the art techniques are presented, followed by a discussion of recent advancements in this exciting field.

**Keywords:** single-molecule spectroscopy, FRET, PIFE, AFM, optical tweezers, RNA folding, RNA helicase, DEAD-box

Published in *RNA Biology* 2013, **10**, 132-147.

## 4.1 Introduction

Originally thought of as simple intermediaries in protein synthesis from a DNA sequence, it has become increasingly clear that RNA molecules play essential roles in almost every aspect of cellular metabolism [36]. RNA functions span diverse processes such as protein synthesis (as mRNA, rRNA, and tRNA) [30, 235, 236, 237, 238], post-transcriptional precursor mRNA processing (as snRNA, and snoRNA) [31, 32, 209, 239, 240, 241], maintenance of telomeres (as telomeric RNA) [242], and control of gene expression (as miRNA, siRNA, aRNA, and riboswitches) [243, 244, 245, 246, 247]. Moreover, RNA molecules are crucial in the action of parasitic entities like retrotransposons, viruses and satellite RNA [248, 249, 250]. Such functional diversity is brought about by the structural diversity of RNA, which can be somewhat surprising considering the limited number of building blocks: the four nucleobases adenine, guanine, cytidine and uracyl [3]. In addition to canonical base pairing, RNA molecules are also capable of forming numerous interactions, such as, base-stacking, non-canonical base-pairing, ribose 2'-hydroxyl bonding and metal ion binding [23, 64]. These interactions are essential to form the native, three-dimensional structures that make RNA molecules function inside living cells [23, 251].

Small functional motifs, such as G-quadruplexes in the untranslated regions of mRNA, are able to form readily [3, 252]. In contrast, long RNA molecules composed of several hundred nucleotides often form stable conformational intermediates that present local minima in the energy landscape. When separated by considerable energy barriers from the next minimum, these intermediates can result in folding kinetic traps [24, 253, 254, 255, 256], which prevent the RNA from readily folding into the more stable native structure [163]. Accumulation of kinetic traps can result in reduced, altered or even lost of functionality, because the native conformation is reached slowly or not at all [257]. As a consequence, kinetic traps must be resolved *in vivo* through the action of helper proteins, such as helicases. Helicases, are ubiquitous enzymes that can catalyse different types of conformational changes and structural rearrangements, going far beyond the traditional view of double strand separating motors [258]. RNA helicases are involved in almost all aspects of cellular metabolism including translation initiation [259, 260], ribosome biogenesis [261, 262], RNA splicing [263, 264], miRNA function [265, 266, 267], RNA transport [258, 268, 269], replication by RNA viruses [270], and unwinding of G-quadruplex structures [271]. RNA helicases are RNA dependent ATPases. Their mechanism can comprise sequences of complex steps involving as many as 75 rate constants in the case of DEAD-box RNA helicases (*vide infra*) [272]. Rapid RNA structural rearrangements occur upon ATP hydrolysis, followed by phosphate and ADP release, thereby turning over to another round of catalysis in which each conformational intermediate displays altered affinity its cognate substrate [273, 274]. Interestingly, dsRNA unwinding is also observed in the absence of ATP, though with slower kinetics and to a lesser extent [273, 274]. Given the vast number of cellular processes RNA helicases are involved in, it is not surprising that individual helicases can strongly differ in translocation rate, directionality, processivity, active versus passive mechanism and step size (i.e. number of base pairs translocated or unwound during each catalytic cycle) [275]. Through phylogenetic analysis, RNA helicases have been grouped and functionally classified into four superfamilies (SF). The most prevalent ones are the structurally similar SF1 and SF2, while a number of viral helicases belong to SF3 and SF4 [276, 277]. The majority of RNA helicases studied to date belongs to the largest subfamily SF2, characterised by ATP-dependent translocation on the nucleic acid substrate and/or induction of RNA conformational changes and further sub-grouped as DEAD-, DEAH- and DExH-box helicases. The subgroup names derive from a highly conserved amino acid motif, in single letter code (e.g., D-E-A-D, Asp-Glu-Ala-Asp), in their helicase domain [275, 276, 278]. Numerous biochemical studies, including gel mobility [279], and unwinding assays [280, 281], have yielded valuable insights into their function, substrate specificity and steady-state kinetics [282]. However, due to ensemble-averaging, these experiments may fail to detect unwinding intermediates and/or conformational or functional heterogeneities within the sample [282]. Through the advent of single-molecule methods, it has become possible to monitor individual macromolecules in real time, thus eliminating population averaging [283, 284, 285]. Rare or transient intermediates can, therefore, be directly observed provided that (i) time resolution is sufficient, and (ii) enough events can be captured to rule out experimental artifacts [122, 161, 284, 286, 287]. Over the years, single molecule approaches have led to an increasing number of fascinating discoveries [141, 169, 170, 288, 289, 290]. Here, we provide an overview on single molecule spectroscopy methods and how they have been successfully employed to characterise the interplay between helicases and their RNA substrates. Technical specifications, advantages and drawbacks associated with individual techniques are also briefly discussed. We would also like to direct the reader to other interesting reviews on related topics [282, 291, 292, 293, 294, 295].

## 4.2 Force-based approaches and their application to study helicase-mediated duplex unwinding

The application of force to single molecules allows for probing and manipulating the folding energy landscape, as well as measuring the force generated by single biomolecules [161, 300]. These approaches do not typically require labeled samples, as they extend the biopolymer between two handles, or attachment sites [161]. Atomic force microscopy (AFM) and optical tweezers have been previously used to study the effect of helicases on RNA substrates (*vide infra*) [301, 302]. Two other force-based techniques are magnetic tweezers and tethered-particle microscopy (TPM), which have both been successfully used to characterise DNA-helicases, but they will not be presented here [303, 304, 305, 306].

### 4.2.1 Atomic force microscopy

Put forward by Binnig and co-workers [307], AFM has become a powerful tool for imaging and manipulating molecules at the nanoscale [308, 309, 310, 311]. At the heart of the mechanical microscope is a flexible cantilever, which acts as a linear spring according to Hooke's law and is used to detect force exerted onto its tip by means of a laser beam that is reflected from the cantilever onto a detector [297]. Typically, the molecule under study is tethered to a surface at one end, and to the AFM tip at the other, such that force can be measured and/or applied (Figure 4.1A) [288, 302], although other experimental approaches have also been used [301]. AFM requires minimal sample preparation and experiments can be carried out in many different environments to obtain data in $nm^3$ spatial and 100 ms temporal resolution [297, 312], offering the possibility of imaging entire proteins and their environment at once (Table 4.1) [297]. The advent of high-speed AFM using smaller cantilevers oscillating at higher frequencies has led

Table 4.1: Single-molecule techniques applied in studying helicase-mediated RNA folding and their technical specifications.

| | Optical tweezers [288, 296] | AFM [288, 297] | smFRET [141, 298] | PIFE [146, 299] |
|---|---|---|---|---|
| Spatial resolution | 0.1 - 2 nm | 0.1 nm in $z$ plane,<br>5 nm in $xy$ plane | 3-8 nm, depending on $R_0$ | below 4 nm |
| Temporal resolution | $10^{-4}$ s | $10^{-5}$ s | confocal: $\geq 50$ µs<br>TIRFM: $\geq 1$ ms | TIRFM: $\geq 1$ ms |
| Force range | 0.1 - 100 pN | 10 - 1000 pN | - | - |
| Typical applications | 3D manipulation, tethered assays, interaction assays | high-force pulling and interaction assays | protein and nucleic acid folding, interaction assays, enzymology | docking/undocking dynamics, protein displacement along RNA |
| Features | low-noise and low-drift dumbbell geometry, no labeling required, *in vivo* measurements (in principle) possible | high-resolution imaging, no labeling, measurements under near-physiological conditions | confocal: observe one molecule at a time TIRFM: image several hundreds of molecules at a time | Image >100 molecules at a time, rather robust against photophysics |
| Limitations | photodamage, sample heating, non-specific | large, high-stiffness probe, large minimal force, nonspecific surface technique | labeling required, photobleaching, dye-sample interaction | labeling, photobleaching, dye-sample interaction |

to time resolutions over 1000 frames/s [313]. However, due to the requirement for physical contact with the sample, its application is currently limited to *in vitro* systems with isolated components. A more detailled description has been given elsewhere [288, 300, 297].

### 4.2.2 Optical tweezers

In the same year that AFM was invented, Chu and co-workers demonstrated particle trapping in a laser beam and suggested its application to study *biological particles* [314]. The technique is based on light-matter interaction; the optical trap is generated by focusing laser light to a diffraction-limited spot using a high numerical aperture objective [288]. In the vicinity of the laser focus, the dielectric object to be trapped experiences a three-dimensional force directed towards the beam waist [288]. Therefore, the laser can be regarded as a microscopic lens that focuses the electro-magnetic field, which causes the particle to move toward the highest intensity point [296]. Particle displacements from the trap center are used to follow and quantify the forces exerted onto a biomolecule, such as an RNA molecule, tethered to the bead. Optical tweezers readily enable measurements of forces as small as 25 fN and sub-nanometer resolution (Table 4.1) [296, 300, 315, 316, 317, 318]. Since their discovery, experiments with optical tweezers have enabled ground-breaking studies of actin-myosin interactions, kinesin motion along microtubules, viral packaging, RNA polymerase transcription, and DNA and RNA folding amongst others [296, 319, 320, 321, 322]. While the use of optical tweezers has become popular due their purely optical nature, they can also present some experimental hindrances such as challenging calibrations [288], local heating resulting from high-intensity, focused laser beams. These issues can complicate the implementation of optical tweezers measurements in live cells, even though the technique has been successfully adapted for *in vivo* studies on other systems [323]. For a more detailled description on optical tweezers, we direct the reader to the following reviews [288, 296, 324].

### 4.2.3 RNA duplex unwinding in the presence of the DEAD-box helicases eIF4A and Ded1

Secondary structure motifs along mRNA molecules can impede the translational machinery search for the start codon required for protein synthesis initiation [15, 325, 326, 327]. Two eukaryotic initiation factors, eIF4A and Ded1 (Table 4.2) [291], belonging to the DEAD-box family (SF2) of RNA helicases, have been shown to resolve RNA secondary structures and unfold RNA hairpins *in vitro* [302, 328, 329]. Using AFM, McCarthy and coworkers studied the mechanism of these two helicases at the single molecule level [302]. The authors used RNA sequences derived from the naturally occurring GCN4 5'-UTR, which were surface-tethered through 5'-end thiolation and attached to the cantilever tip using biotinylated nucleotides in the poly-A tail (Figure 4.1A). The RNA was then stretched by pulling the cantilever away from the surface. The force exerted on the RNA was determined by measuring the bending of the cantilever with the reflected laser beam (Figure 4.1A). The resulting force-extension curves revealed the force necessary to stretch and refold the RNA at a certain distance (Figure 4.1B, C). Each molecule could be pulled ∼20-30 stretching/refolding cycles before dissociation of the RNA, backbone cleavage or disruption of the biotin/streptavidin bond. Incorporation of a stable 25 GC basepair stem-loop has been shown to inhibit translational initiation in *S. cerevisiae* [327]. In the AFM experiments, incorporation of the same hairpin resulted in an abrupt drop in force in the stretching curve (arrow, Figure 4.1B), which was absent in the refolding curve. This drop was assigned to the opening of the hairpin [327]. In the presence of saturating Ded1, the force required to open the hairpins decreases from ∼150 pN to ∼90 pN (Figure 4.1C). Comparison between the two helicases showed that Ded1 is more effective in reducing the force necessary to unfold the hairpin, even in the presence of elF4B, an elF4A cofactor that enhances its unwinding rate. In addition, the unwinding activity of eIF4A/B saturates well before the stem loop is fully unwound. Based on these results, the authors concluded that Ded1 is a more efficient facilitator of stem-loop unwinding than elF4A/B, and likely the major unwinding factor on natural mRNA substrates. Based on earlier bulk studies [329], the authors proposed that eIF4A acts via an ATP-dependent steric mechanism, whereby the helicase binds to single-stranded RNA adjacent to the stem loop structure, thus causing its partial unwinding, whereas Ded1 dynamically unwinds the stem loop through low processive, ATP-dependent translocation [302].

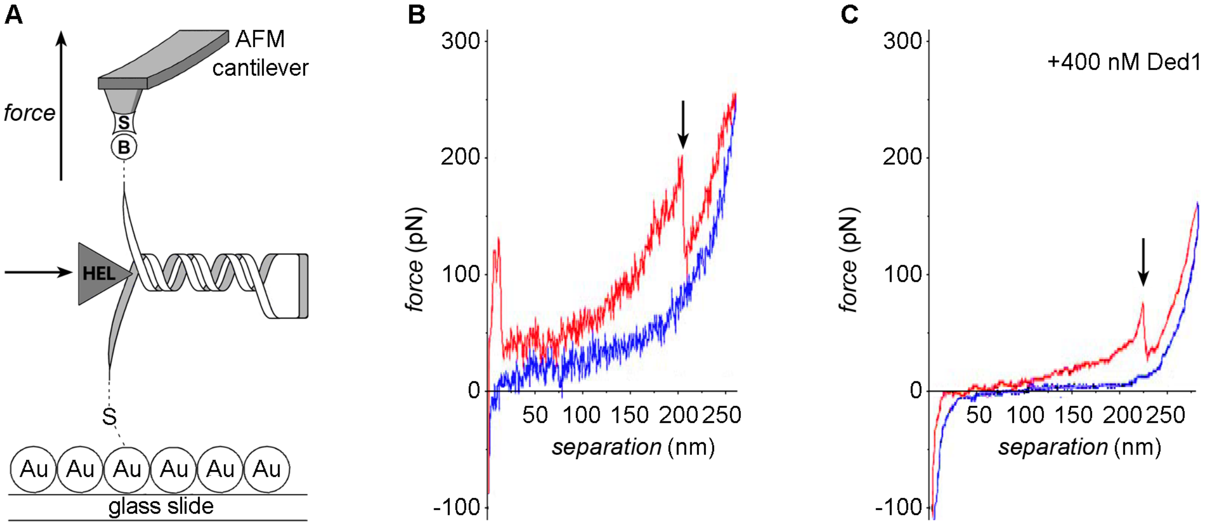


Figure 4.1: Helicase induced RNA hairpin unfolding using AFM force spectroscopy. (A) RNA immobilization involves tethering the 5'-end to the gold-coated surface (*Au*) and attaching the biotinylated 3'-end (*B*) to a streptavidin-coated (*S*) silicon nitride ($Si_3Ni_4$) tip. Force-extension curves were recorded in the presence and absence of eIF4A or Ded1, represented as "HEL". (B, C) Representative force-extension plots representing stretching of a single RNA molecule containing a GC-rich stem-loop. (C) Stretching of the same RNA in the presence of 400 nM Ded1. The pull curve (red) runs from left to right and the approach curve (blue) from right to left. The arrow indicates the force and extension when the hairpin unfolds. In the presence of Ded1, the hairpin unfolds at lower force. Adapted from [302].

In this study, the relative efficiency of two eukaryotic helicases to separate was assessed from the decrease in unwinding force. Future studies may involve eIF4G, which has been shown to stimulate eIF4A via a conformational guidance mechanism [330]. A valuable extension of analogous studies will involve time dependent force-extension plots (*vide infra*). This will allow finding key mechanistic features of processive helicases, such as the number of base pairs unwound/ATP hydrolysed, the physical step size (mean distance/translocation step), as well as the kinetic step size (mean number of base pairs unwound/rate-limiting step) [282].

### 4.2.4 RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP

The hepatitis C virus (HCV) non-structural protein 3 (NS3) is an SF2 DEAH-box helicase that plays a key role in viral RNA replication and an important drug target against HCV infection (Table 4.2) [331, 332]. In a study by Dumont and co-workers, NS3-catalysed unwinding of dsRNA and translocation long the substrate was characterise using optical tweezers [333]. A 60 bp RNA hairpin with handles was attached to two beads and held under constant strain (Figure 4.2A). The end-to-end distance of the RNA was measured in the presence and absence of NS3 to monitor helicase-induced RNA unwinding. Force-extension curves show that the presence of NS3 decreased the force required for unfolding from 20.4 pN (Figure 4.2B, green) to 5 pN, though, this depends on the GC-content of the sequence [334]. To monitor NS3 unwinding activity, the RNA substrate was held at constant force in the 5 to 17 pN range, in the presence of the helicase (Figure 4.2B, red). The resulting time trajectories revealed stepwise increases of inter-bead distance, which was subsequently converted into basepairs (Figure 4.2C), and interpreted as unwinding steps and pauses [335]. A histogram of pairwise distances (Figure 4.2D) revealed a periodic pattern of steps, suggesting that the cyclic behavior of NS3 is coordinated by ATP in discrete steps of 11 ± 3 bp that include rapid smaller substeps of 3.6 ± 1.3 bp, which reflect actual unwinding. However, in a follow-up study, step size was shown to be decreased to 9 ± 2 bp in the presence of long GC-rich stretches that were absent in the substrate used in the initial study [334]. Interestingly, the authors also observed apparent backward steps of 12 ± 3.6 bp corresponding to the stepwise refolding of the substrate after unwinding. This pronounced processivity is somewhat surprising, given that NS3 differs in only one of six conserved motifs from its phylogenetically-close relatives from the DEAD-box family of helicases, which are almost non-processive [258].

Mean pause duration of the NS3 helicase was found to be inversely proportional to ATP concentration, indicating that exiting from pauses requires ATP hydrolysis. Analysis of pause distributions below and above the Michaelis-Menten constant ($K_m$) pointed the authors to propose a kinetic mechanism by which the helicase exits from pauses in two steps, only one of which involves ATP hydrolysis. The occurrence and duration of stalling was shown to depend upon the presence of GC islands of six or more basepairs but not three [334]. Furthermore, the stepping velocity of NS3 between pauses increases with ATP concentration indicating that each step consists of ATP-dependent substeps. Based on the kinetic data, the authors concluded that NS3 must bind ATP in each substep during unwinding [333]. The processivity of NS3 did not significantly depend upon either ATP or NS3 concentration above $K_D$, but was strongly dependent on the force applied and also found to be affected by the base composition of the duplex [334]. In contrast, pausing and stepping velocity were independent of force indicating that NS3-mediated RNA unwinding activity is not limited by strand separation but by translocation. Although the stepping velocity and $K_m$ were in good agreement with prior bulk experiments, the single molecule assays were seemingly in conflict with previous ensemble experiments reporting a step size of 18 ± 2 bp as compared to 11 ± 3 bp [338]. The authors proposed that NS3 dimerization in bulk experiments may account for this difference, while the 3'-overhang

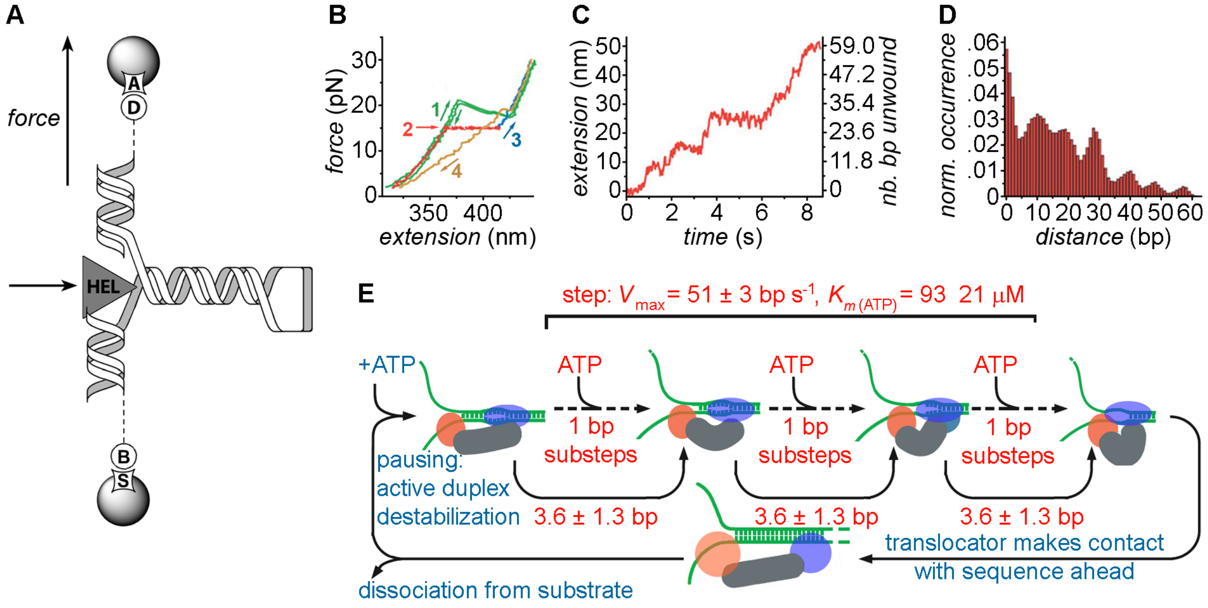


Figure 4.2: Revealing NS3 translocation and unwinding mechanism by optical tweezers. (A) A 60 bp RNA hairpin is flanked by two RNA/DNA duplexes. The bottom duplex (599 bp) is attached to a bead via biotin-streptavidin linkage (*B*, *S*). The top duplex (535 bp) is attached to an anti-digoxigenin coated bead via digoxigenin (*A*, *D*). NS3-mediated (*HEL*) duplex unwinding is followed upon application of external force. (B) Unwinding experimental steps: mechanical folding and unfolding of substrate in the absence of NS3 shows a transition at 20.4 pN, the force required for hairpin unwinding (green), the presence of NS3 decreases the force required for duplex unwinding (red), force is brought to 30 pN to fully extend the nucleic acid strand (blue). The RNA hairpin is allowed to refold at 2 pN force (yellow). (C) Extension of bead separation during RNA unwinding by NS3 reveals discrete steps and pausing. (D) A pairwise distance histogram of the unwinding trace shown in (C) is later subjected to Fourrier analysis to determine the apparent unwinding step size. (E) Unified model for NS3 helicase activity. Nucleic acid substrate binding is followed by ATP-dependent destabilization duplex by the translocator domain (blue ellipse) [334]. The RNA hairpin is unwound by helix-opening domain (orange circle) in fast ATP-dependent 1 bp-substeps adding up to steps of 3.6 ± 1.3 bp, unwinding 7 to 13 basepairs per apparent step depending on the GC content [333, 334, 336, 337]. Subsequently, the translocator moves forward to start a new catalytic cycle [333]. At this stage, GC-rich sequences increase the probability of helicase dissociation [334]. Adapted from [333, 334].

used in the single-molecule study did not allow for dimer formation [333]. This hypothesis was further confirmed in subsequent bulk experiments [334]. In the proposed model, the NS3 monomer has two RNA binding sites; one site acts as a translocator and the other as a helix opener [333]. Both translocation and unwinding take place as ATP-dependent substeps, resulting in 11-basepair unwinding bursts. Following results further indicated that these substeps are preceded by active destabilization of the double-stranded region to be unwound, as opposed to opportunistic forward motion of the enzyme upon thermal fraying of the duplex [334]. In an independent fluorescence-based study carried out with double-stranded DNA, Myong *et al.* proposed an alternative model analogous to a loaded spring involving three successive ATP hydrolysis events driving the two domains forward in one-base-pair steps, followed by an abrupt 3 bp-separation using DNA [336]. To reconcile these findings, Cheng and co-workers have recently used a homopolymeric GC sequence to avoid sequence-dependent variation in step size and to slow down the helicase [337]. Hence, their experiments at very low ATP concentrations (to further slow down the helicase) achieve 1 bp resolution [337]. They observed both large 11-bp unwinding steps and 1-bp unwinding events, thus reconciling fluorescence experiments with force spectroscopy studies. However, they also observed statistically significant higher-order 1.5-, 2- and 2.5-bp unwinding events [337]. As the rate constants for unwinding were independent of substepping size, the authors proposed that higher-order steps are due to release of nucleotides held in previous rounds of unwinding, in agreement with recent crystallographic data (Figure 4.2E) [337, 339]. This example illustrates how different single molecule approaches can be used alongside to improve our understanding of helicase-mediated double-stranded RNA unwinding.

## 4.3 Fluorescence-based methods and their applications in elucidating RNA-helicase interactions

### 4.3.1 Single-molecule Förster Resonance Energy Transfer (smFRET)

FRET is a process by which energy is non-radiatively transferred between two dipoles, typically the transition dipole moments of two fluorophores that are referred to as donor and acceptor. Efficient energy transfer requires the emission spectrum of the donor to overlap with the excitation spectrum of the acceptor, which satisfies the resonant (equal energy) condition. Energy transfer efficiency depends very strongly on the distance between the two dipoles ($R$), and is calculated as $E = (1 + (R/R_0)^6)^{-1}$, where $R_0$ is the Förster radius at which half of the energy is transferred. The Förster radius is typically in the 30-80 Å range, but depends on the spectral properties of the fluorophores, their relative orientation and their local environment [141]. Owing to its distance-dependence, FRET is regularly referred to as a *molecular ruler* to probe inter- and intramolecular distances in biological systems [278, 340]. Experimentally, the lack of information on relative fluorophore orientation makes it challenging to determine the exact energy transfer efficiency [341]. Therefore, the apparent energy transfer efficiency is often simply calculated as $FRET = I_A/(I_A + I_D)$, where $I_D$ and $I_A$ are the fluorescent intensities of the donor and acceptor, respectively.

smFRET is a popular and adaptable approach to study biomolecular mechanisms of isolated molecules because FRET provides the sensitivity and selectivity necessary to detect single molecules dynamics in real time (Table 4.1) [141, 342]. It has, therefore, been applied to study both the interaction of surface-immobilised RNAs with a helicase [340, 205, 343], and to probe helicase conformational dynamics in solution [344, 345, 346, 347]. By monitoring individual molecules, smFRET can help reveal sample heterogeneity and the presence of transient intermediates [189].

### 4.3.2 Protein-induced fluorescence enhancement (PIFE)

An important limitation in the interpretation of smFRET data arises from fluorophore blinking (intersystem crossing to long-lived, non-fluorescent triplet states) and photobleaching (light-induced chemically modified non-fluorescent adducts) [146]. Despite recent advancements in the suppression of such photophysics through addition of reducing (or oxidizing) agents [150, 151, 348], dark states, blinking and unstable emission still present a challenge in single-molecule fluorescence microscopy [189]. An interesting solution to this problem is the use of protein induced fluorescence enhancement (PIFE), which requires only one label [299, 349].

Cyanine dyes, which consist of two aromatic moieties linked by a polymethylene chain, can adopt *cis* and *trans* conformations, but the latter conformer exhibits considerably higher fluorescence quantum yield [152]. Therefore, *cis-trans* isomerisation decreases the apparent quantum yield of cyanine dyes and their fluorescence lifetime [152]. The principle behind PIFE relies on a local increase of the viscosity of the fluorophore environment due to the proximity of a protein, resulting in hindrance of the *cis-trans* isomerisation and hence, altered fluorophore quantum yield. This allows to directly monitor protein binding and dissociation dynamics [349, 299, 350]. However, in contrast to the ratiometric smFRET method that minimises the effect of inherent emission fluctuations of single fluorophores [183], PIFE data is prone to produce noisy data. Because specific environmental changes in the vicinity of fluorophores are challenging to predict, interpretation of experimental data requires careful calibration and well-designed control experiments [299].

### 4.3.3 Helicase-assisted secondary structure conversion towards and against thermodynamic equilibrium

Helicases are multifunctional enzymes that, not only separate double-stranded nucleic acids, but they can also drive structural rearrangements required for function [351, 352]. Interestingly, stable secondary RNA structures have been observed to be converted into thermodynamically less stable ones [351]. Shortly after the single molecule characterisation of Ded1-catalysed unwinding of RNA duplexes [302], Jankowsky and co-workers used smFRET in conjunction with bulk biophysical methods to characterise Ded1-catalysed RNA structure remodeling (Table 4.2) [353]. The authors studied whether Ded1 catalyses conversion of a stable RNA duplex into a thermodynamically less stable one by first disassembling existing duplex structures, or through recruitment of multiple strands to form a tripartite intermediate (Figure 4.3A). To address this question, they designed a three-strand RNA substrate such that two, mutually exclusive, duplex structures of distinct stabilities could be formed (Figure 4.3A, left and right). One of the strands contained a biotin label for surface-tethering, the second strand contained the donor fluorophore (Cy3) and the third strand contained the acceptor fluorophore Cy5 (Figure 4.3A). Therefore, the more stable duplex is characterised by Cy3 emission alone (*i.e.*, zero FRET), the least stable duplex is characterised by the loss of the Cy3 strand (and thus, Cy3 fluorescence), and a possible tripartite complex by the presence of both Cy3 and Cy5 in close proximity (*i.e.*, emission of both fluorophores via FRET). Having determined the equilibrium ratio of the two duplex structures under different conditions, the authors observed structural interconversion against thermodynamic equilibrium upon addition of Ded1 and ATP using non-denaturing gel electrophoresis. A gel mobility shift was accompanied by the generation of single-stranded RNA. In the absence of ATP, strand exchange was also observed, but to a lesser extent, and single-stranded RNA was not detected.

In single-molecule FRET experiments, the resulting time trajectories clearly showed the reversible formation of a transient tripartite intermediate complex (Figures 4.3B and C, black arrows). In rare instances, the authors also observed the whole remodeling pathway with initial formation of the tripartite intermediate followed by dissociation of the Cy3 labeled strand (Figure 4.3C, red arrow). The authors concluded that Ded1-assisted structure conversion can proceed via two pathways: (i) ATP-dependent strand exchange involving complete dissociation of the RNA strands, followed by protein-facilitated annealing; (ii) An ATP-independent pathway promoting the formation of a tripartite intermediate complex (Figure 4.3A) [353].

These results confirmed previous studies in which Ded1 was shown to be a low processive, ATP-dependent helicase [302, 343]. The characterisation and identification of the transient, tripartite intermediate and the remodeling pathway would not have been possible using solely ensemble-averaged techniques, underscoring the strength of single-molecule techniques to resolve transient intermediates in the reaction pathway. The authors thereby revealed an RNA remodeling activity of the Ded1 helicase in addition to its previously known duplex unwinding activity. However, a limitation in this assay is that formation of the less stable duplex, which is characterised by the loss of the Cy3 strand, and thus, loss of both Cy3 and Cy5 fluorescence, could not be distinguished from Cy3 photobleaching. One possible alternative approach is to invert the labeling scheme, because formation of the tripartite complex would result in a high FRET signal followed by loss of Cy5 fluorescence, indicating strand exchange. An elegant way to rule out photophysical artifacts is by using alternating laser excitation (ALEX), which consists of exciting both fluorophores alternatively, thereby determined both FRET and fluorophore stoichiometry which unambiguously reports on photobleaching and blinking [354, 355].

### 4.3.4 Unwinding initiation by the viral RNA helicase NPH-II

NPH-II is a viral SF2 helicase with NTP-dependent unwinding activity, which plays an important role in transcription termination and viral RNA export [356, 357, 358]. Experiments in vitro have shown that NPH-II unwinding initiation can be two-orders of magnitude slower than strand separation, thus making initiation the rate limiting step [358]. Interestingly, the enzyme can even hydrolyse hundreds of ATP molecules before unwinding [358]. This makes NPH-II an interesting model system to study unwinding initiation. After their initial smFRET study with Ded1, Jankowsky and co-workers characterised NPH-II-mediated dsRNA unwinding initiation using single-molecule FRET. The authors surface-immobilised a Cy3-labeled dsRNA with a Cy5-labeled 3'-single-stranded overhang (Figure 4.4A). The overhang is required *in vitro* for

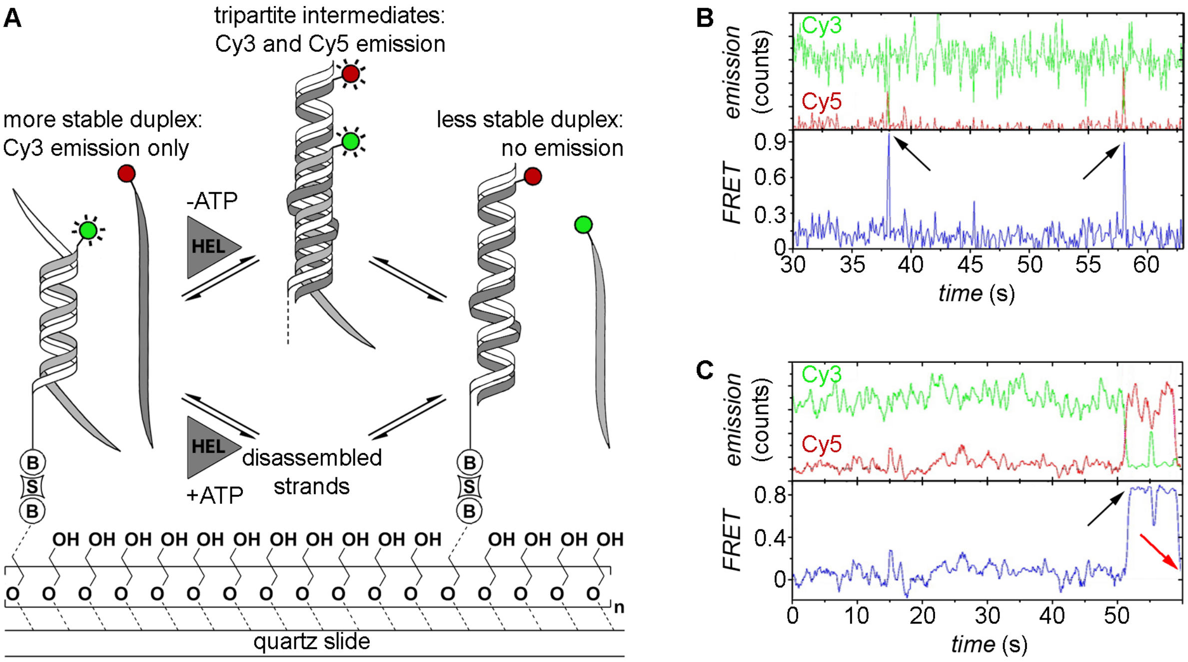


Figure 4.3: Ded1-mediated RNA remodeling monitored by smFRET. (A) Experimental design and results. The Cy3-labeled RNA strand (light grey, Cy3: green circle) forms a thermodynamically stable duplex with the immobilization strand (white) that is surface-tethered via a biotin-streptavidin linkage (*B*, *S*), while the Cy5-labeled strand (dark grey, Cy5: red circle) is in solution. Strand exchange in the presence of Ded1 (*HEL*) and in the absence of ATP proceeds via a tripartite intermediate characterised by a sudden increase in Cy5 emission though FRET, followed by the formation of the thermodynamically less stable duplex accompanied by a complete loss of fluorophore emission (upper pathway). In the presence of both Ded1 and ATP, strand exchange involves complete disassembly of the more stable duplex and subsequent formation of the less stable one (lower pathway). Passivation of the quartz slide is achieved through coating with polyethylene glycol to prevent non-specific binding of the protein to the slide.161 (B-C) Representative fluorophore emission and FRET time traces. Cy3 and Cy5 emission over time (green and red traces) displays FRET-typical anticorrelated behavior (upper plots). FRET over time reveals the formation of the tripartite intermediate, as indicated by sudden burst of FRET (lower plots, black arrows), followed either by dissociation of the Cy5-labeled strand (B) or strand exchange and complete absence of fluorophore emission (C, red arrow). Figure modified from [353].

NPH-II recruitment [279, 356, 358, 359]. Single-molecule time trajectories recorded under various experimental conditions (Figure 4.4B) were used to build time-binned FRET histograms that report on the conformational distribution of the 3'-end overhang (Figures 4.4C-H) [170].

In the absence of helicase, the free RNA yielded a distribution centered around 0.85 FRET (Figure 4.4C), as expected for the two fluorophores in close proximity (Figure 4.4A). Addition

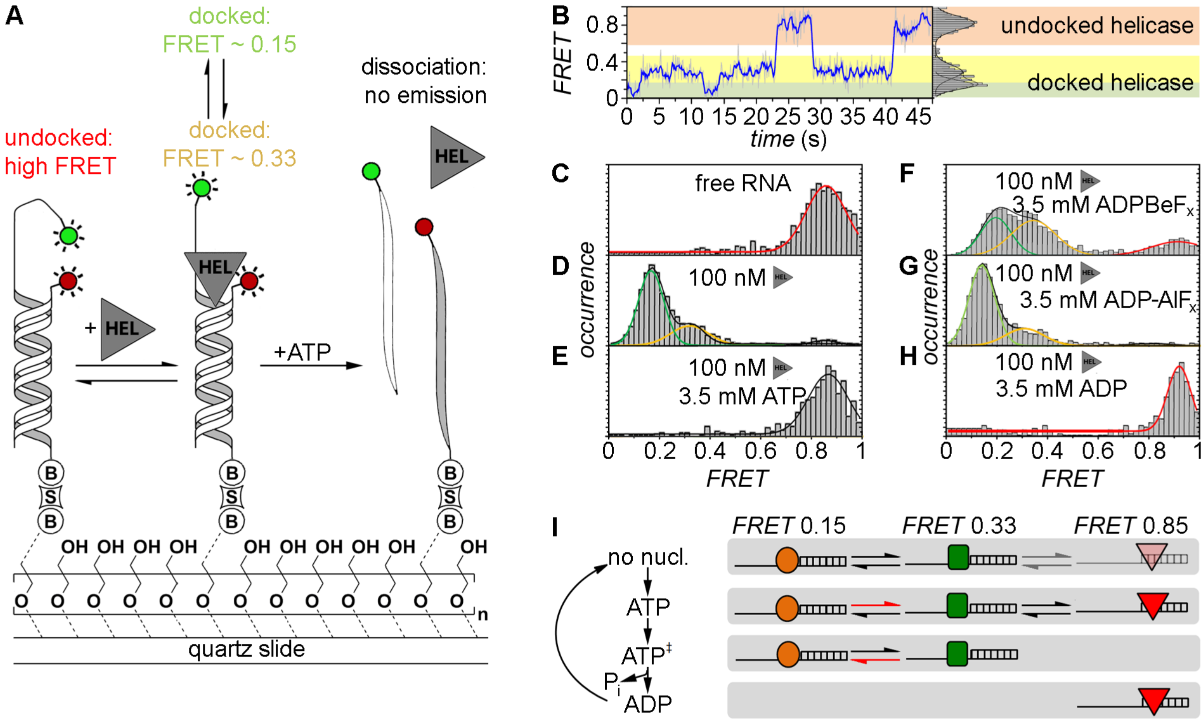


Figure 4.4: Characterizing NPH-II-catalysed RNA duplex unwinding. (A) Experimental design and results. A fluorophore-labeled 19-bp RNA duplex (Cy3: green circle, Cy5: red circle) with a 24-nt 3' extension was immobilised on the PEG-passivated quartz slide via a biotin-streptavidin linkage (*B*, *S*).161 The initial high FRET value of 0.85 is diminished in response to NPH-II binding (*HEL*) and fluctuates between two discrete low FRET values, indicating and increased inter-dye distance. Addition of ATP triggers duplex unwinding ultimately leading to complete loss of emission upon strand separation. (B) Representative smFRET trajectory (1 nM NPH-II, no ATP) showing transition between the helicase-unbound state (FRET 0.85, highlighted in orange), and the helicase-bound states (FRET 0.15 and 0.33, highlighted in green and yellow, respectively). (C-H) Averaged FRET histograms, each built from over 100 individual FRET time traces. Imaging conditions: (C) only RNA, (D) RNA and 100 nM NPH-II, (E) 100 nM NPH-II, 3.5 mM ATP, (F) 100 nM NPH-II, 3.5 mM ADPBeFx (a ground-state analogue), (G) 100 nM NPH-II, 3.5 mM ADP-AlFx (a transition state analogue), (H) 100 nM NPH-II, 3.5 mM ADP. (I) Basic model for unwinding initiation by NPH-II relying on altered substrate affinities along the ATP hydrolysis cycle. Without nucleotide, NPH-II binds both ssRNA and dsRNA and the NPH-II-ssRNA complex readily alternates between two conformations. ATP binding impedes dissociation from ssRNA and changes the kinetics of bound state transitions. In the ATP transition state, NPH-II no longer binds to dsRNA and interconversion kinetics change again. The helicase associates with dsRNA upon ATP hydrolysis and phosphate dissociation. Different shapes mark the different conformational states if NPH-II traversed during unwinding initiation. Figure adapted from [340].

of NPH-II alone induced a conformational change in the RNA-NPH-II complex that led to two distinct conformations with FRET distributions centered at ∼0.15 and ∼0.33 (Figure 4.4D). Addition of ATP shifted the FRET distribution back to 0.85 (Figure 4.4E). However, the time trajectories clearly revealed a rapid decrease in FRET followed by a loss of fluorescence, a signature indicative of helicase binding followed by duplex unwinding. To further test the mechanism, the authors replaced ATP by non-hydrolysable analogues blocking unwinding at different states of ATP hydrolysis [360, 361, 362, 363, 364, 365]. In the presence of ADP-$BeF_X$ (a ground state analogue), all three FRET states were observed (Figure 4.4F). Addition of ADP-$AlF_X$ (a hydrolysis transition state analogue) resulted in population of the two low FRET states (Figure 4.4G). And in the presence of the hydrolysis product, ADP, only the high FRET peak was observed (Figure 4.4H).

Based on these results and other ensemble-averaged control experiments, the authors proposed a two-state kinetic model for unwinding initiation, and explained it through alternating ssRNA/dsRNA binding properties of the protein at different stages of the ATPase cycle (Figure 4.4H): Without nucleotide, the helicase binds to both ssRNA and dsRNA, displaying intrinsic dynamics (*i.e.*, 0.15 and 0.33 FRET states). Initial ATP binding stabilises the bound complex preventing helicase dissociation from the ssRNA. In the hydrolysis transition state, NPH-II no longer binds the duplex. Following ATP hydrolysis, the helicase no longer binds to ssRNA. Thus, single-molecule experiments provide a mechanistic explanation for the maintenance of RNA-helicase contacts during duplex unwinding initiation, a long-standing question in the field [358, 366, 367]. A future challenge will be to clarify why a high number of ATP molecules are often hydrolysed prior to strand separation and which specific molecular events induce strand unwinding. This could be achieved by directly labeling the helicase in order to probe its conformational dynamics prior to unwinding, thus complementing smFRET experiments using fluorophore-labeled RNA substrates.

### 4.3.5 Single-molecule analysis of Mss116-mediated group II intron folding

The *Saccharomyces cerevisiae* DEAD-box helicase Mss116 is essential for mitochondrial group I and group II intron splicing *in vivo* (Table 4.2) [152]. Mss116 activity has been extensively studied in ensemble-averaged experiments using the D135 ribozyme, which is a minimal model system for the *S. cerevisiae* group II intron ai5$\gamma$ [79, 368]. While Mss116 helicase activity depends on ATP binding and hydrolysis, the Mss116-mediated D135 folding mechanism is still not fully understood. As an ideal tool to dissect RNA folding pathways, smFRET experiments were performed to characterise Mss116-mediated D135 folding [205].

Fluorescently-labeled D135 ribozymes were surface-immobilised onto PEG-coated quartz slides via a biotin-streptavidin linkage [369, 370], and FRET was measured using total internal reflection fluorescence (TIRF) microscopy (Figure 4.5A) [204]. Previous smFRET studies have shown that, in the absence of Mss116, D135 requires high $Mg^{2+}$ concentrations for efficient folding in vitro [204]. Moreover, the presence of $Ca^{2+}$ altered the ribozyme's structural dynamics and impeded its function [87]. Building averaged FRET histograms from over 100 individual time trajectories recorded under precedented high-salt conditions suggested the presence of three distinct FRET states (Figure 4.5B) [204]. As group II introns fold in an $Mg^{2+}$ ion-dependent fashion [371], it was rationalised in an earlier smFRET study conducted at different $Mg^{2+}$ ion concentrations that the three FRET states correspond to an extended folding intermediate (I, ∼0.1 FRET), a folded intermediate (F, ∼0.4 FRET) and the native state (N, ∼0.6 FRET) [204]. Assignment of the native state was further supported by experiments conducted in the presence of the cleavable substrate 17/7, which significantly increased the population of N [204]. smFRET trajectories confirmed that F is an obligatory folding intermediate in the D135 folding pathway because only a small fraction of molecules ($< 2$ %) folded directly from I to N. Under near-physiological conditions (Figure 4.5C), smFRET experiments in the absence

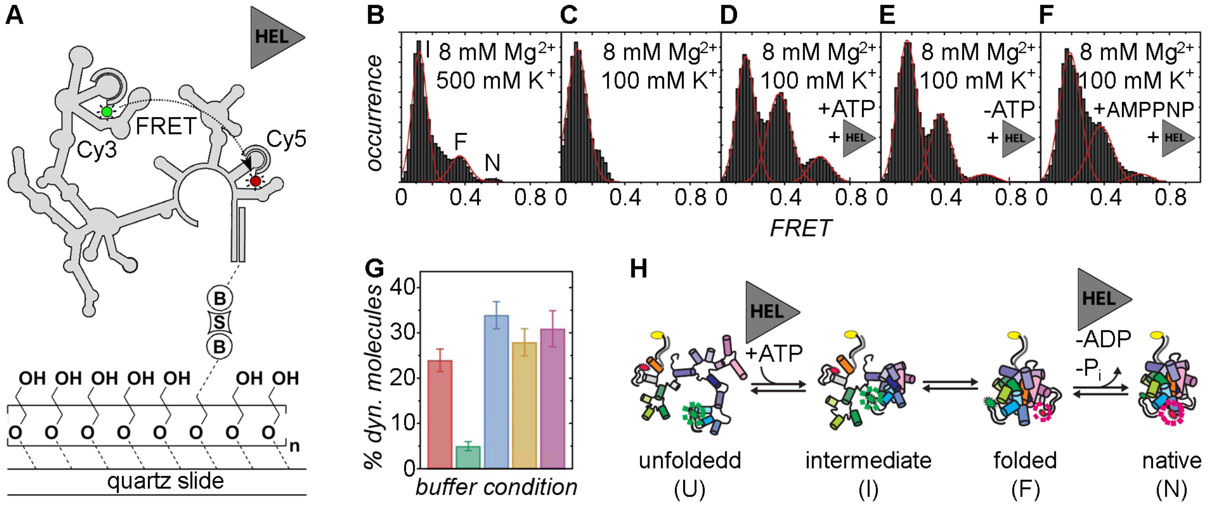


Figure 4.5: Mss116-mediated group II intron folding using smFRET. (A) Experimental design. The fluorophore-labeled D135 ribozyme (Cy3: green cycle, Cy5: red cycle) is immobilised on a PEG-coated quartz slide via a biotin-streptavidin linkage (*B*, *S*). Structural interconversion under different folding conditions is monitored by following FRET efficiency over time. (B-F) Averaged FRET histograms, each built from over 100 single-molecule time traces. Imaging conditions: (B) 8 mM $Mg^{2+}$, 500 mM $K^{+}$. Three FRET distributions are observed, termed *intermediate* (*I*), *folded* (*F*) and *native* (*N*) based on earlier results [204, 87]. (C) 8 mM $Mg^{2+}$, 100 mM $K^{+}$. Only the *I* FRET state is observed at near-physiological conditions. (D) 8 mM $Mg^{2+}$, 100 mM $K^{+}$, 25 nM Mss116, 1 mM ATP. Addition of Mss116 and ATP shifts the distribution of FRET states towards the folded intermediate and the native state. (E-F) 8 mM $Mg^{2+}$, 100 mM $K^{+}$, 25 nM Mss116 (and 1 mM AMPPNP). Effect of ATP hydrolysis on D135 folding. Prevalence of the native state is lowered in the absence of ATP (E) and in the presence of non-hydrolysable AMPPNP (F). (G) Percentage of dynamic molecules at different imaging conditions. red: 500 mM $K^{+}$, 8 mM $Mg^{2+}$; green: 8 mM $Mg^{2+}$, 100 mM $K^{+}$; blue: 8 mM $Mg^{2+}$, 100 mM $K^{+}$, 25 nM Mss116, 1 mM ATP; yellow: 8 mM $Mg^{2+}$, 100 mM $K^{+}$, 25 nM Mss116; purple: 8 mM $Mg^{2+}$, 100 mM $K^{+}$, 25 nM Mss116, 1 mM AMPPNP. (H) Proposed model of Mss116-mediated group II intron folding. D135 interconverts between four conformations (*unfolded* (*U*), *I*, *F*, *N*). Mss116 promotes the transition from *U* to *I*, even in the absence of ATP. It further catalyses ATP-dependent conversion from *F* to *N*. Figure modified from [205].

of Mss116 showed only the presence of the extended intermediate (I), confirming that the ribozyme alone cannot form the native state in low salt. Addition of Mss116 and ATP, however, restored all three FRET states (Figure 4.5D), confirming that Mss116 promotes the formation of the native state. Control experiments with other, basic RNA binding proteins showed that electrostatic interactions alone account for the initial stabilization of the folded intermediate but not the native state. Similar experiments in absence of ATP (Figure 4.5E) showed that Mss116 alone can induce the formation of all three FRET states, albeit structural conversion is promoted to a lesser extent. Similar results were obtained for the non-hydrolysable ATP analog AMPPNP (Figure 4.5F). Since Mss116 can facilitate formation of N both in the presence and absence of ATP, the authors proposed that ATP hydrolysis may play a role in Mss116 dissociation and recycling to increase helicase-assisted folding efficiency.

As an inherent advantage, smFRET provides a direct means to characterise the ribozyme's dynamics and to identify the presence of subpopulations. Single-molecule time traces revealed the presence of two subpopulations: static molecules that did not undergo any changes in FRET during the observation time (minutes) and dynamic molecules that displayed FRET transitions. In the absence of Mss116 and under high salt conditions, the dynamic subpopulation represented ∼25 % of the observed molecules (Figure 4.5G, red). Decreasing the ionic strength to near physiological levels, decreased the dynamic population to below 5 % (Figure 4.5G, green). Under these ionic conditions, addition of Mss116 and ATP increased the dynamic population to ∼35 % (Figure 4.5G, blue). Removal of ATP, or the presence of the non-hydrolysable ATP analogue (AMPPNP) also yielded ∼25-30 % of dynamic molecules. The large subpopulations of static molecules in the I state in the absence of Mss116 indicate the presence of at least one high-energy activation barrier for folding between I and F.

Traditionally, folding rate constants ($k_1$, $k_{-1}$, $k_2$, $k_{-2}$) are obtained by measuring dwell times in each FRET states using manual thresholds. But noisy datasets or species with close FRET values can make this analysis challenging [284]. A powerful alternative is the use of hidden Markov Models (HMM) to analyse the smFRET trajectories, and to determine the folding rate constants and characterise the energy barriers on the folding pathway (Figure 4.5H). In the absence of Mss116, high salt was found to accelerate the folding reaction by ∼3-fold, while slowing down the unfolding reaction by ∼2-fold. Interestingly, addition of Mss116 and ATP had a similar effect. These results indicate that Mss116 functions by decreasing the activation barriers between folding intermediates and by stabilizing the native state.

In summary, this study showed that the ribozyme folds through a series of obligatory intermediates and that Mss116 functions by facilitating formation of specific species (Figure 4.5H): formation of the folded intermediate is promoted by electrostatic interactions, even in the presence of non-specific basic proteins. However, efficient transition to the native state requires the presence of both Mss116 and ATP. In contrast to other RNA-helicase studies *in singulo*, the authors used a native Mss116 substrate (a large, catalytically competent group II intron ribozyme), under near-physiological ionic conditions. Such large substrates can leads to complex data analysis. For example, the free rotation assumption of the cyanine dyes may not hold true. However, the authors carried out anisotropy measurements to address this issue [204]. Alternatively, linearly polarised nanosecond ALEX can help monitor fluorophore rotational freedom in real time [355].

### 4.3.6 Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA

During viral infections, the Retinoid acid Inducible-Gene protein (RIG-I) specifically recognises 5'-triphosphate on viral RNA and subsequently initiates the antiviral immune response in the host (Figure 4.6 and Table 4.2) [372, 373]. RIG-I has a modular structure comprising two N-terminal caspase activation and recruitment domains (CARDs), a central ATPase domain, required for antiviral signaling, and a C-terminal regulatory domain (RD) [374, 375]. In a recent study, Myong and co-workers employed smPIFE to characterise the role of the central domain and to address whether RIG-I action is primarily induced by the presence of a 5'-triphosphate or the prevalence of dsRNA, both of which are characteristic markers of viral RNA [349]. They assessed ability of the full length protein (wtRIG), as well as two truncated versions of RIG-I (RIGh and svRIG) to recognise different DY547-labeled RNA substrates, followed by translocation along a stretch of dsRNA (Figures 4.6A-C) [349].
In an initial set of experiments, fluorophore-labeled dsRNA devoid of a 5'-triphosphate was surface-immobilised onto a polymer-passivated quartz slide (Figure 4.6A) [349]. RIGh binding to the dsRNA resulted in sharp fluctuations of DY547 fluorescence in the single molecule trajectories (Figure 4.6D) that became more frequent in the presence of both RIGh and ATP [349]. To test whether RIGh translocates during upon substrate binding, the authors measured the distribution of dwell times between fluorescence bursts using 25 and 40 basepair dsRNA. As expected for a translocating enzyme, the distribution of dwell times in the bound state increased for the substrate with the longer end-to-end distance (Figure 4.6E), indicating that RIGh unidirectionally translocates on dsRNA. Analogous experiments in the presence of wtRIG and ATP revealed similar, though less frequent fluctuations in fluorescence trajectories. The resulting dwell time histograms (i) confirmed the trend towards longer dwell times in the presence of the 40bp-substrate and (ii) revealed a much broader distribution with ∼15-fold higher average value as compared to RIGh (Figure 4.6F). Decreased translocation velocity along dsRNA lacking the 5'-triphosphate in the presence of CARDs is in agreement with earlier reports, and was explained by inhibition of ATP hydrolysis by conducting further ATP-dependent experiments [376]. A splice variant of RIG-I was constructed by removing amino acids 36-80 in the first CARD domain (svRIG, Figure 4.6C). The single molecule PIFE trajectories and the resulting dwell time histograms (Figure 4.6C) revealed that svRIG behaves similar to RIGh, indicating that translocation activity was restored for svRIG. These results confirm that both CARD domains of RIG-I are required to suppress translocation activity in the absence of the 5'-triphosphate [376].

Next, the authors evaluated the effect of a 5’ triphosphate attached to a single-stranded overhang on RIG-I translocation (Figure 4.6B). Single-molecule PIFE trajectories and the resulting dwell time histograms show that wtRIG translocates ∼20-fold faster on the 5’-triphosphate substrate than on standard dsRNA. This suggests that the 5’-triphosphate prevents ATPase inhibition by the two CARDs, and the authors proposed that RIG-I translocation is preceded by 5’-triphosphate recognition. To rule out scanning on the single stranded overhang prior to translocation along the dsRNA, the authors varied the ratio of ssRNA and dsRNA stretches

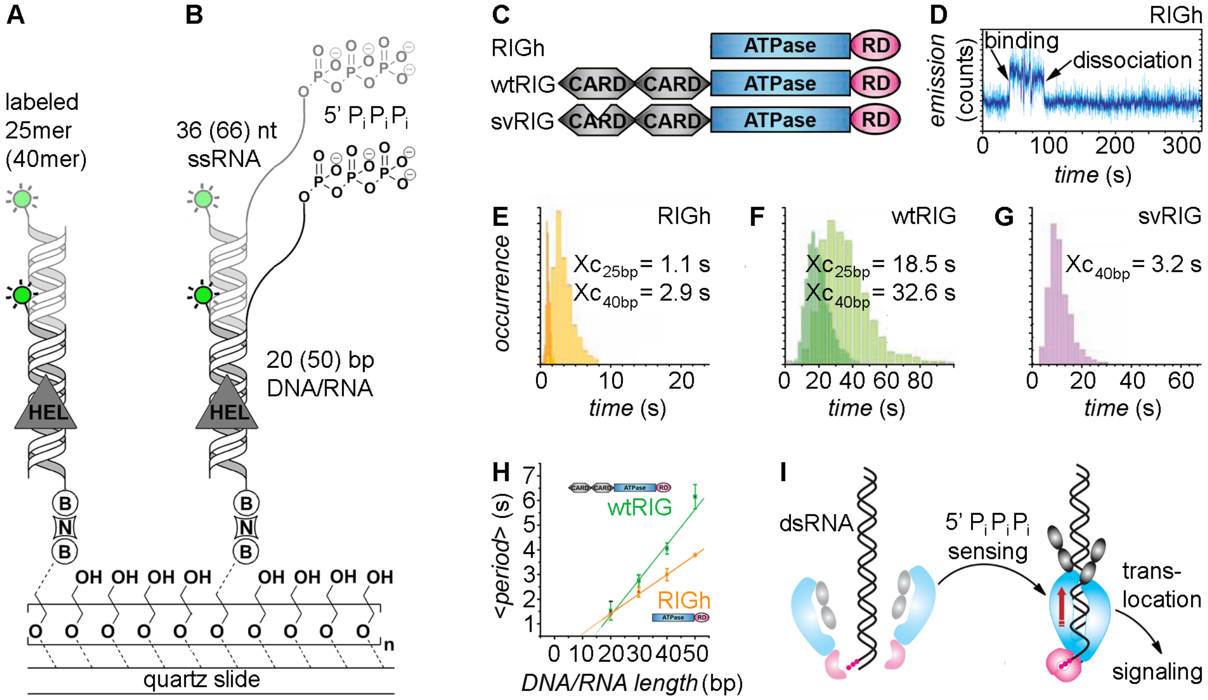


Figure 4.6: Probing RIG-I translocation on dsRNA using PIFE. (A-B) Experimental design. A 25/40bp dsRNA with blunt ends (A) or a 20-50bp dsRNA with a 66-36 nt ssRNA overhang and a 5’ triphosphate (“P$_i$P$_i$P$_i$”) (B) is labeled with a single DY547 fluorophore (green circle) and immobilised on a PEG-passivated surface via biotin-neutravidin (“B”, “N”) [370]. (C) Depiction of three modular RIG-I variants used in this study. RIGh consists of the central DExH-box ATPase domain and a C-terminal regulatory domain (“RD”). wtRIG additionally has two N-terminal caspase activation and recruitment domains (“CARD”). In svRIG, one of the CARDs is non-functional. (D) Representative time trajectory recorded in the presence of RIGh and the blunt end RNA substrate shown in (A). Helicase binding is accompanied by a sudden increase of fluorophore emission. (E-G) Dwell-time analyses for time traces recorded in the presence of RIGh (E), wtRIG (F), and svRIG (G) for 25-bp and 40-bp blunt end dsRNA shown in (A). RIGh and svRIG translocate faster along dsRNA than wtRIG. In all cases, the average time required for end-to-end translocation increases with the substrate length. (H) RIG-I translocation on dsRNA in the presence of 5’-triphosphate, average time spent in the bound state versus duplex length. (I) Proposed model for pathogen-associated molecular pattern (“PAMP”) signal integration by RIG-I.168 Binding of the RIG-I regulatory domain (pink) to RNA 5’ triphosphates induces RIG-I dimerization as decribed previously.174 This triggers the translocase domain (blue), followed by translocation along the dsRNA substrate (red arrow) and induction of a CARD signaling conformation (gray). Figure modified from [349].

and observed linear increase of dwell times with increasing length of the duplex for both RIGh and wtRIG (Figure 4.6H). These results suggest a model where RIG-I first recognises the 5'-triphosphate of viral RNA, which in turn prevents ATPase inhibition by the two CARDs, thus enabling RIG-I unidirectional translocation and the immune response cascade (Figure 4.6I). However, the way in which the immune response is linked to the unidirectional translocation of RIG-I remains unclear. The authors speculated that repetitive shuttling might inhibit binding of pathogenic proteins playing a role in the viral replication cycle. This does not explain why ATPase-deficient RIG-I mutants capable of unidirectional translocation do not elicit an immune response *in vivo*. Future studies should be carried out with native RNA substrates, as these may form higher-order structures that can be recognised by RIG-I thus modulating its activity, and in the presence of viral RNA-binding proteins known to play a role in replication.

Table 4.2: Some RNA-helicases studied by single-molecule approaches, in order of appearance in text.

| Helicase | Full name (www.uniprot.org) | Subgroup | Biological role(s) | Substrate(s) | Method | Ref. |
|---|---|---|---|---|---|---|
| eIF4A | eukaryotic initiation factor 4A | DEAD-box | translation initiation | dsRNA | AFM, smFRET | [302]<br>[353] |
| Ded1 | ATP-dependent RNA helicase DED1 | DEAD-box | translation initiation | ssRNA, dsRNA | AFM, smFRET | [302]<br>[353] |
| HCV NS3 | Hepatitis C virus non-structural protein 3 | DEAH-box | viral replication | dsRNA/DNA with 3' ss overhang [279, 332] | optical tweezers, smFRET* | [333]<br>[334]<br>[336]*<br>[337] |
| NPH-II | Nucleoside triphosphate phosphohydrolase II | DEAD-box | transcription termination, RNA export [356] | viral dsRNA | smFRET | [340] |
| Mss116 | ATP-dependent RNA helicase MSS116, mitochondrial | DEAD-box | splicing | group I and II introns | smFRET | [205] |
| RIG-I | retinoic acid inducible gene I | DExD/H-box | RNAi, immune response [377, 375] | viral dsRNA | PIFE | [349] |
| YxiN | ATP-dependent RNA helicase YxiN | DEAD-box | ribosome biogenesis [378] | 23S rRNA [379] | smFRET | [344]<br>[345]<br>[346]<br>[347] |

* study performed with dsDNA

## 4.4 Conclusion

It is increasingly clear that RNA helicases play numerous crucial and diverse roles in essentially all aspects of RNA metabolism. Here, we have reviewed recent single-molecule studies that have provided unprecedented information on helicase mechanisms. Force-based methods have allowed for the observation of individual mechanistic cycles of different helicases with near-basepair resolution, unveiling unwinding forces and processivity. In turn, fluorescence-based methods have been used to observe and characterise helicase-mediated strand exchange, helicase-RNA interaction, helicase conformational dynamics, as well as helicase translocation along a double helix in real time. The resulting models did not only expand our understanding of helicase mechanisms, but they also underpinned the mechanical and functional diversity between different helicases that are highly related with regard to the conservation of specific sequential motifs. Such detailled characterisation would have been hardly accessible with conventional ensemble-averaged techniques, making single-molecule spectroscopy a powerful tool to study these key biologically reactions.

In turn, some of the studies reviewed make use of artificially designed RNA model systems (such as RNA hairpins) which may simplify data interpretation, but my also fair to reveal the biological function of the enzyme. Other aspects mimicking a living cell were also missing. Therefore, future studies should not only include other biologically important RNA helicases such as DICER, but also known protein cofactors, for example eIF4G that has been demonstrated to increase eIF4A activity, as well as proteins known to bind to the natural RNA substrate, such as viral cofactors in the case of RIG-I. A possibly interesting approach to simulate the crowded environments inside living cells would be imaging in the presence of molecular crowding agents that have been successfully employed in other research fields [380]. Methodological efforts have already resolved mysteries associated with single-molecule data (see Section 4.2.4). We believe that technical advances will continue to contribute to making single-molecule spectroscopy more robust. Current efforts aim at the development of multidimensional FRET to simultaneously map multiple domains, a new generation of fluorophores with improved photostability and higher spatial and temporal resolution [125, 146, 286, 300, 381, 382]. We anticipate that such technical improvements will ultimately lead to imaging folding of single RNA molecules inside live cells, which will provide further fascinating insights into the function of RNA helicases.

# Chapter 5

# Metal ion induced kinetic heterogeneity of the intron-exon recognition in single group II introns

Danny Kowerko,[1,*] Sebastian L.B. König,[1,*] Miriam Skilandat,[1] Daniela Kruschel,[1] Lucia Cardo,[2] Roland K.O. Sigel[1]

1 Institute of Inorganic Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; 2 Institute of Biomedical Research, University of Birmingham, United Kingdon

**ABSTRACT:** RNA is commonly believed to undergo a number of sequential folding steps prior to reaching its functional fold, being the global minimum in the free energy landscape. However, it is increasingly clear that several functional conformations are often in coexistence corresponding to multiple (local) minima in the folding landscape. Here, we use the 5' exon-intron recognition duplex of a self-splicing ribozyme as a model system to study the influence of the physiologically important cations $Mg^{2+}$ and $Ca^{2+}$ on tertiary RNA structure formation. Single-molecule FRET performed in conjunction with biochemical bulk experiments reveal that near-physiological $Mg^{2+}$ and $Ca^{2+}$ concentrations strongly promote interstrand association. Moreover, the presence of divalent metal ions leads to pronounced heterogeneity in RNA-RNA interaction kinetics, suggesting the presence of multiple active conformations. Using NMR, we locate specific $Mg^{2+}$ binding pockets close to the RNA-RNA interaction sites by NMR. This unprecedented combination of NMR and smFRET demonstrates for the first time that a rugged free energy landscape coincides with incomplete occupation of specific metal ion binding sites under near-physiological cation concentrations. Unconventional kinetics associated with RNA folding are therefore likely to originate from a spectrum of conformations that differ with regard to the occupation of metal ion binding sites.

* Equal contributions. Submitted to *Proc. Natl. Acad. Sci. USA* in December 2013.

## 5.1 Introduction

RNA-RNA interactions, *i.e.* the formation of secondary and tertiary structure, are key to functional RNA molecules. Tertiary interactions are thereby structurally much more diverse and complicated and they constitute the central part of any higher order architecture. The folding pathway of any RNA depends on the exact consecutive formation of a series of secondary and tertiary contacts. As both the folding pathway and the final structure correlate directly with function, there is a fundamental interest in characterizing the basic principles of RNA-RNA interactions on every level. To form higher order structures, the electrostatic repulsion created by the negative charges throughout the phosphate-sugar backbone must be overcome by positive charges. An estimated 10-20 % of the negative charge is compensated by metal ions that bind specifically, both with regard to the ion species involved as well as to the binding site [2]. As a consequence, cations play a crucial role in global RNA structure and functionality [51, 383].

Self-splicing group II introns are abundant in nature and among the largest catalytic RNAs. Moreover, they are considered ancestors of the eukaryotic spliceosome, telomeres, and retrotransposons [384]. One example is the group II intron Sc.ai5$\gamma$ from the yeast mitochondrial *cox1* gene, in which site specificity of the two sequential transesterfication reactions is ensured by proper base pairing between the distal exon-binding sites (5' cleavage: EBS1 and 2; 3' cleavage: EBS3) and their cognate intron-binding sites (IBS1, 2, 3) [51]. The intron folding pathway and catalysis is strictly dependent on $Mg^{2+}$ and is perturbed by small amounts of $Ca^{2+}$ [87, 204, 385]. Interestingly, $Ca^{2+}$ can reach low mM concentrations in mitochondria, and $Ca^{2+}$ and the *cox1* gene encoded cytochrome oxidase are involved in apoptosis [386].

Single-molecule Förster resonance energy transfer (smFRET) is ideally suited to study the conformational behavior of biological macromolecules in response to the presence of different metal ion identities. According to the principle of ergodicity, such single molecule experiments provide all information of the molecular ensemble [166]. In addition, data on single molecule folding also reveal static and dynamic heterogeneity within the sample, unveil individual conformations and minor species along the folding pathway, and they allow for quantitative characterisation of the associated kinetics. This is important, because the vast conformational space sampled by biomolecules often leads to the occurrence of kinetic traps and/or multiple native states, resulting in differences between individual molecules that were folded and measured side-by side under identical conditions [161]. Such heterogeneity has been precedented for a number of RNA molecules, including group I introns [219, 224], the hairpin ribozyme [195, 197, 198, 199, 200], and the cyclic-di-GMP riboswitch [213], The possible origins mostly remain in the dark and the quantitative characterisation is challenging [164, 387].

Here, we use the 5' exon-intron recognition site EBS1/IBS1 in the Sc.ai5$\gamma$ ribozyme as a basic system to study both thermodynamic and kinetic aspects of $Mg^{2+}$- and $Ca^{2+}$-mediated RNA-RNA tertiary interaction using smFRET. Additionally, nuclear magnetic resonance (NMR) is applied to characterise specific $M^{2+}$ binding. The here used so-called d3'EBS1*/IBS1* has two AU base pairs replaced by GCs to ensure proper duplex formation. This system has been validated previously to be fully active in the ribozyme and its NMR structure has been solved recently [50, 388]. For FRET experiments, the 5'-ends of the d3'EBS1* hairpin and the seven nucleotide long IBS1* are labeled with the fluorophore pair Cy3 and Cy5 (Figure 5.1A). Cy3-d3'EBS1* further carries a biotin at the 3'-end for subsequent immobilisation on the quartz slide as described previously to be imaged with a TIRF microscope [290].

## 5.2 Materials and Methods

### 5.2.1 Chemicals and RNA

All chemicals used for the preparation of buffers and stock solutions were at least *puriss p.a.* and were purchased from Sigma-Aldrich AG (Buchs, Switzerland). Buffers and solutions were prepared using double-distilled autoclaved water (ddH$_2$O) and were subjected to additional filtering using 0.2 µm syringe filters immediately prior to use. RNA sequences for smFRET experiments were purchased from Microsynth AG (Balgach, Switzerland) and purified by denaturing gel electrophoresis (18 % w/v acrylamide/bis-acrylamide, running buffer 1x TBE) and RP-HPLC (X-TERRA C8 semi-preparative column, from 5 to 25 % v/v acetonitrile in 100 mM TEEA, pH 7.0, in 30 minutes). Sequences used in NMR experiments were prepared by *in vitro* transcription as described [50].

### 5.2.2 Native gel electrophoresis

20 pmol Cy3-d3'EBS1*-biotin and an increasing quantity of Cy5- IBS1* (0.25 to 5 equivalents) dissolved in 10 µL running buffer (5 mM MOPS, 2 mM NaOH, 50 µM MgCl$_2$ or 50 µM CaCl$_2$, pH 6.90) were loaded onto a polyacrylamide gel (18 % w/v acrylamide/bis-acrylamide,

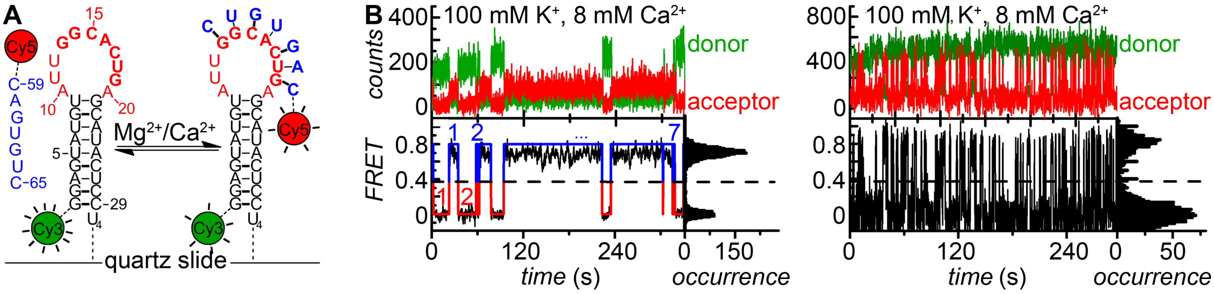


Figure 5.1: Experimental design and representative smFRET time traces. (A) Docking of Cy5-IBS1* onto the surface-immobilised Cy3-d3'EBS1* is characterised by a concomitant appearance of Cy5 emission due to FRET, which allows to monitor the influence of M$^{2+}$ on docking/undocking. (B) Representative fluorescence intensity and *FRET* time traces. Top: Corrected donor and acceptor emission over time. Bottom: *FRET* efficiency time traces reveal fluctuations between the docked state (high *FRET*) and the undocked state (zero *FRET*) at various rates. Thresholding at *FRET* = 0.375 (dashed line) was used to extract dwell times in the zero *FRET* state ($t_{zero,i}$, red) and the high *FRET* state ($t_{high,i}$, blue).

dissolved in running buffer). The gel was run at 4 °C in running buffer until the fragments had migrated to the lower half of the gel. The resulting bands were recorded with a Typhoon 9410 multipurpose scanner (GE Healthcare, Glattbrugg, Switzerland).

### 5.2.3 UV melting experiments

Temperature dependent absorption was recorded at 260 nm on a Cary 500 Scan UV/vis spectrophotometer (Varian Inc., Palo Alto, CA, USA) equipped with a Cary temperature controller in quartz cuvettes with 1 cm path length (Hellma GmbH & Co. KG, Müllheim, Germany). Samples contained 3 µM Cy3-d3'EBS1*-biotin/Cy5-IBS1* dissolved in standard buffer containing 0-8 mM $MgCl_2$/$CaCl_2$ and were degassed and overlaid with paraffin oil prior to measurements (Table 5.1). Absorption was monitored in a temperature range of 80-10 °C at a temperature changing rate of 0.5 °C/min, followed by van't Hoff analysis as described [17, 389].

Table 5.1: Solutions used throughout smFRET and UV thermal melting experiments.

| Solution name | Composition | Ref. |
|---|---|---|
| T50 buffer | 50 mM Tris-HCl, 50 mM NaCl, pH 7.50 | [170] |
| Biotinylated BSA solution | 1 mg/ml, dissolved in T50 buffer | [170] |
| Streptavidin solution | 50 µg/mL, dissolved in T50 buffer | - |
| Standard buffer | 50 mM MOPS, 100 $K^+$, $Ca^{2+}$ or $Mg^{2+}$ at varying concentration, pH 6.90 | - |
| d3'EBS1* solution | 10 pM, dissolved in standard buffer | - |
| 100x oxygen scavenger solution (OSS) | 80 µL T50 buffer, 2170 U/mL catalase, 165 U/mL glucose oxidase | [111] |
| Trolox stock solution | Trolox in standard buffer, pH 6.90 | [150] |
| Reaction buffer | 25 nM IBS1*, 1 mM Trolox (dilution from stock), 1x OSS (dilution from stock), 1 % (w/v) D-glucose, dissolved in standard buffer | - |

### 5.2.4 Single molecule FRET experiments

d3'EBS1* was immobilised within a self-built microfluidic channel using a biotin-streptavidin linker according to a protocol modified from [170] and [290] (for a description of the buffers and solution please refer to Table 5.1): The chamber was flushed with 200 µL T50 buffer to remove impurities from the surface, followed by the injection of 50 µL biotinylated BSA solution, and incubation for 10 minutes. Then, the channel was flushed with 200 µl T50 buffer and 50 µL streptavidin solution, followed by 5 minutes incubation. Excess streptavidin was removed by washing with 100 µL standard buffer and the channel was flushed with 80 µL d3'EBS1* solution and incubated for 5 minutes. Finally, 200 µL reaction buffer containing IBS1*, an enzymatic oxygen scavenger solution (OSS; glucose oxidase: Sigma-Aldrich AG, Buchs, Switzerland; catalase: Thermo Scientific, Rockford, IL, USA) to reduce photobleaching as well as 1 mM Trolox to suppress dye blinking were added and measurements were started [111, 150]. All smFRET experiments were performed at room temperature upon excitation at 532 nm. Both donor and acceptor emission intensities were monitored over 400 seconds using a total internal reflection fluorescence (TIRF) microscope equipped with a charge-coupled device (10 frames/s, 4000 frames) [170, 290].

### 5.2.5 Movie processing and data analysis

Single fluorophore emission time traces were manually selected for anticorrelated donor/acceptor emission using Matlab-based software SIRA (SIf Reader and Analyser) [390]. Single-molecule clips were imported, followed by identification of single fluorophore emission, co-localisation of the corresponding donor and the acceptor coordinates and creation of time traces. The resulting raw time traces were corrected for background noise using an aperture photometry approach, photo-bleaching, leakage of donor emission into the acceptor channel and direct acceptor excitation [125, 189]. Smoothing was performed in the absence of short dwell times in the docked/undocked state using a non-linear forward-backward filter [391, 392]. Corrected donor and acceptor emission intensities, $I_{\mathrm{D}}^{\mathrm{D}}$ and $I_{\mathrm{A}}^{\mathrm{D}}$, were used to calculate the apparent *FRET* efficiency as described [175]:

$$FRET = \frac{I_{\mathrm{A}}^{\mathrm{D}}}{I_{\mathrm{D}}^{D} + I_{\mathrm{A}}^{\mathrm{D}}} \tag{5.1}$$

For each buffer condition, more than 250 *FRET* efficiency time traces of 400 seconds observation time matching the selection criteria were evaluated and categorised into the following subgroups [213]:

- Type I: molecules undergoing at least two transitions, referred to as dynamic molecules.
- Type II: molecules undergoing one transition.
- Type III: molecules undergoing no transition, referred to as static molecules.

  - Type IIIa: molecules remaining in the zero *FRET* state.
  - Type IIIb: molecules remaining in the high *FRET* state.

Cross-sample variability was estimated via bootstrapping using the freely available software package BOBA FRET [166].

### 5.2.6 Single-molecule control experiments

Alternating laser excitation (ALEX) at 532 nm and 640 nm was performed using mechanical shutters synchronised with the CCD read-out as described [125]. This extension yields an additional photon count over time, *i.e.* acceptor intensity upon direct acceptor excitation $I_{\mathrm{A}}^{\mathrm{A}}$, which allows to calculate the stoichiometry of fluorophores and the detection of photophysical artifacts [28, 354, 393].

$$FRET = \frac{I_{\mathrm{D}}^{\mathrm{D}} + I_{\mathrm{A}}^{\mathrm{D}}}{I_{\mathrm{D}}^{\mathrm{D}} + I_{\mathrm{A}}^{\mathrm{D}} + I_{\mathrm{A}}^{\mathrm{A}}} \tag{5.2}$$

Additionally, the influence of $Mg^{2+}$ and $Ca^{2+}$ on the photophysics of Cy3 and Cy5 was characterised. For this purpose, Cy3-d3'EBS1* or Cy5-d3'EBS1* were surface-immobilised as described above, followed by monitoring $I_{\mathrm{D}}^{\mathrm{D}}$ and $I_{\mathrm{A}}^{\mathrm{A}}$ at 0, 1, 10, or 100 mM $Mg^{2+}$/$Ca^{2+}$ over 400 seconds. Care was taken to reach photon counts similar to the intensity levels observed in smFRET experiments. Subsequently, photostability, fluorophore bleaching, blinking, and the occurrence of long-lived dark states were quantified as described [149].

### 5.2.7 Bulk fluorescence control experiments

Bulk fluorescence experiments were performed with a Safire 2 microplate reader (Tecan Group Ltd., Männedorf, Switzerland). 3995-COR 96-well plates (Vitaris, Baar, Switzerland) display minimal autofluorescence and crosstalk and are coated to prevent surface attachment of the sample. Samples contained sulfonated cyanine dye NHS esters (Cy3 or Cy5, GE Healthcare, Glattbrugg, Switzerland) or sulfonated cyanine dyes tagged to RNA (d3'EBS1* or IBS1*) dissolved in reaction buffer at a consistent concentration of 100 nM (Table 5.1). The sample also contained 0, 1, 10, 100 or 1000 mM $Mg(NO_3)_2$ or $Ca(NO_3)_2$. Each measurement was repeated using three independent samples to calculate the standard deviation. MilliQ dd$H_2O$, 1x OSS in MilliQ $H_2O$, as well as reaction buffer with and without OSS were also measured to determine the contributions of individual compounds to the background noise. Fluorescence intensities were quantified performing fixed wavelength scans at constant equal slit width and $\lambda_{\mathrm{exc/em}}$ = 530/560 nm for Cy3 and $\lambda_{\mathrm{exc/em}}$ = 600/660 nm for Cy5, respectively. Fluorescence spectra were recorded, averaged and smoothed with a sliding average filter using a window of 10 data points [284], followed by normalisation to 1. Anisotropy $r$ was quantified as

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{5.3}$$

where $I_{\parallel}$ and $I_{\perp}$ refer to the emission intensities when the emission polariser is parallel or perpendicular relative to the direction of polarised excitation [109].

### 5.2.8 Association constants and Gibbs free energy

Association and dissociation constants ($K_\mathrm{A}$, $K_\mathrm{D}$) were determined either from the binned apparent *FRET* efficiency histograms according to Equation 5.4 or by analysis of dwell-times in the docked/undocked state according to Equation 5.5. The latter were determined either by thresholding or Hidden Markov modelling as described [195, 214, 394].

$$K_\mathrm{A} = \frac{1}{K_\mathrm{D}} = \frac{c(\mathrm{DA})}{c(\mathrm{D})c(\mathrm{A})} = \frac{c(\mathrm{d3'EBS1^*/IBS1^*})}{c(\mathrm{d3'EBS1^*})c(\mathrm{IBS1^*})} \simeq \frac{p(FRET_\mathrm{high})}{p(FRET_\mathrm{zero})[\mathrm{IBS1^*}]_\mathrm{total}} \tag{5.4}$$

$$K_\mathrm{A} = \frac{1}{K_\mathrm{D}} = \frac{1 - \frac{t_\mathrm{zero}}{(t_\mathrm{zero}+t_\mathrm{high})}}{\frac{t_\mathrm{zero}}{(t_\mathrm{zero}+t_\mathrm{high})}[\mathrm{IBS1^*}]} \simeq \frac{t_\mathrm{high}}{t_\mathrm{zero}[\mathrm{IBS1^*}]_\mathrm{total}} \tag{5.5}$$

$c$(D), $c$(A), and $c$(DA) are the absolute concentrations of undocked *FRET* donor D, free *FRET* acceptor A, and donor-acceptor complex in the docked state DA. The relative population of docked and undocked d3'EBS1* molecules ($c$(d3'EBS1*/IBS1*) and $c$(d3'EBS1*)) is determined from the binned apparent *FRET* efficiency histograms p(FRET) shown in Figures 5.2 and 5.3. $p(FRET_\mathrm{high})$ are the cumulative counts with $FRET > 0.375$, $p(FRET_\mathrm{high}) = \int_{0.375}^{1.200} p(FRET)\mathrm{d}FRET$, corresponding to the docked species. The population of undocked species $p(FRET_\mathrm{zero})$ is defined by $FRET < 0.375$, with $p(FRET_\mathrm{high}) = \int_{-0.200}^{0.375} p(FRET)\mathrm{d}FRET$. The undocked fraction $P$(<0.375) is thus defined as the ratio of undocked molecules to the sum of docked and undocked molecules:

$$\frac{t_\mathrm{undocked}}{t_\mathrm{undocked} + t_\mathrm{docked}} = \frac{\int_{-0.200}^{0.375} p(FRET)\mathrm{d}FRET}{\int_{-0.200}^{1.200} p(FRET)\mathrm{d}FRET} = P(< 0.375) \tag{5.6}$$

Accordingly, the docked fraction is $P$(>0.375) = 1 - $P$(<0.375). Integration of the FRET histograms $p$(*FRET*) and normalisation to 1 yields the so-called cumulative probability distribution $P$ [284]. $\mathrm{t_{zero}}$ and $\mathrm{t_{high}}$ are the dwell times in the zero and high *FRET* state observed under a given buffer condition. The concentration of undocked IBS1* ($c$(IBS1*)) is approximated by the total concentration $c_\mathrm{total}$(IBS1*), assuming that $c_\mathrm{total}$(IBS1*) >> $c$(d3'EBS1*).
The $K_\mathrm{A}$ value of an individual molecule $n$ was determined according to Equation 5.7:

$$K_{\mathrm{A},n} = \frac{1}{K_{\mathrm{D},n}} \simeq \frac{\langle t_{\mathrm{high},n}\rangle}{\langle t_{\mathrm{zero},n}\rangle c_\mathrm{total}(\mathrm{IBS1^*})} \tag{5.7}$$

with

$$\langle t_{\text{zero},n} \rangle = \frac{\sum\limits_{i=1}^{i_{\max}} t_{\text{zero},i,n}}{i_{\max}} \tag{5.8}$$

and

$$\langle t_{\text{high},n} \rangle = \frac{\sum\limits_{j=1}^{j_{\max}} t_{\text{high},j,n}}{j_{\max}} \tag{5.9}$$

$\langle t_{\text{zero},n} \rangle$ and $\langle t_{\text{high},n} \rangle$ are the mean dwell-times averaged over $i_{\max}$ and $j_{\max}$ dwell-times of a molecule $n$. Note that $i_{\max}$ and $j_{\max}$ vary among different molecules and $|i_{\max}\ j_{\max}| = 1$ or 0. Averaged $K_A$ values from single molecule FRET traces were calculated from the sum of all dwell times of the total number of analysed molecules $n_{\max}$.

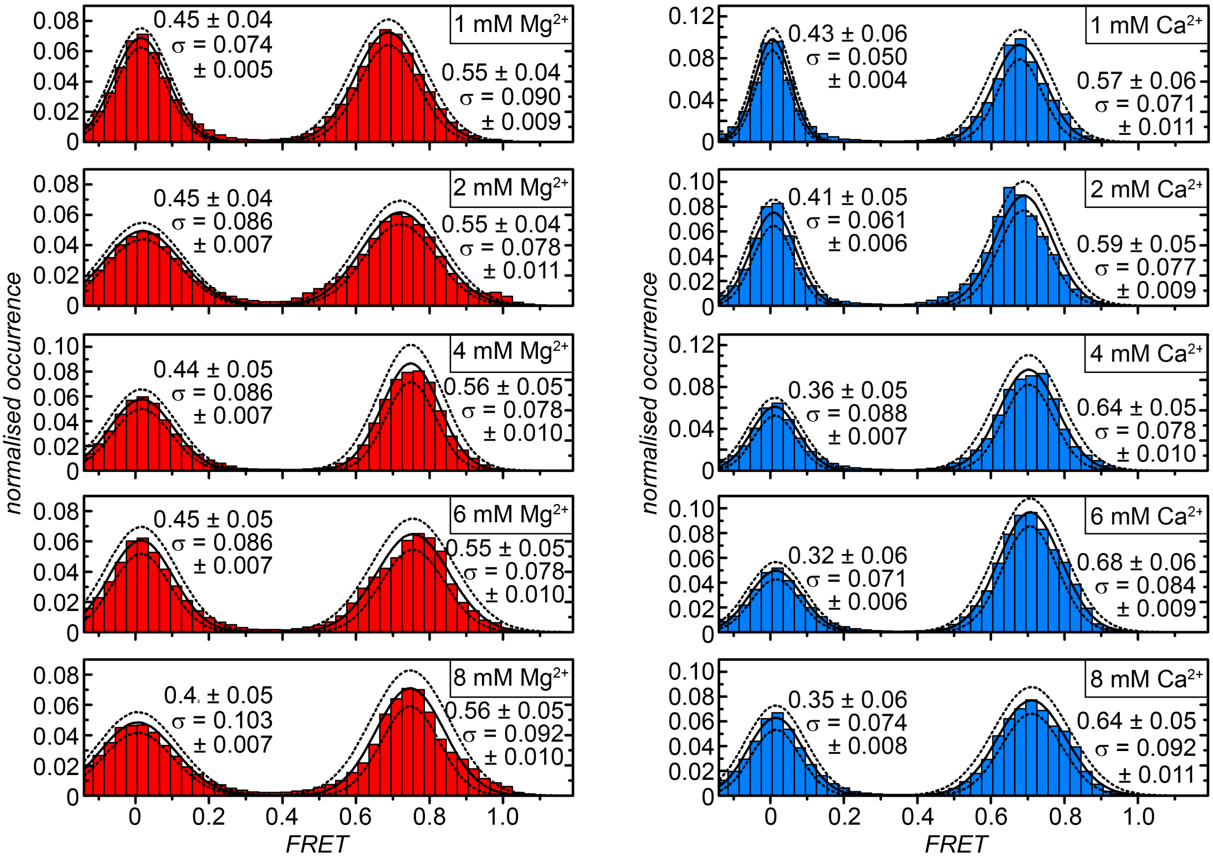


Figure 5.2: Normalised averaged histograms built from smFRET time traces recorded at different $Mg^{2+}$ (left) or $Ca^{2+}$ concentrations (right). The zero *FRET* distribution corresponds to undocked d3'EBS1*, while the high *FRET* peak corresponds to an IBS1* molecule docked onto d3'EBS1*. Values for bound and unbound fractions, and errors ($2\sigma$ determined by bootstrapping and regression using BOBA FRET (13). Imaging conditions: 50 mM MOPS, 100 mM KCl, 25 nM IBS1*, 1 % (w/V) D-Glucose, 1x OSS, 1 mM Trolox, pH 6.90, divalent metal ions as indicated.

$$K_A = \frac{1}{K_D} \simeq \frac{t^{\text{sum}}_{\text{high}}}{t^{\text{sum}}_{\text{zero}} c_{\text{total}}(\text{IBS1}^*)} \tag{5.10}$$

$$\langle t_{\text{high},n} \rangle = \frac{\sum_{j=1}^{j_{\max}} t_{\text{high},j,n}}{j_{\max}} \tag{5.11}$$

The Gibbs Free Energy $\Delta G^\circ$ of single molecules was then calculated from $K_A$ according to

$$\Delta G^0_n = \mathrm{k_B} lnK_{D,n} = -\mathrm{k_B} T\, lnK_{A,n} \tag{5.12}$$

with $\mathrm{k_B}$ equals the Boltzmann constant and $T$ the absolute temperature.

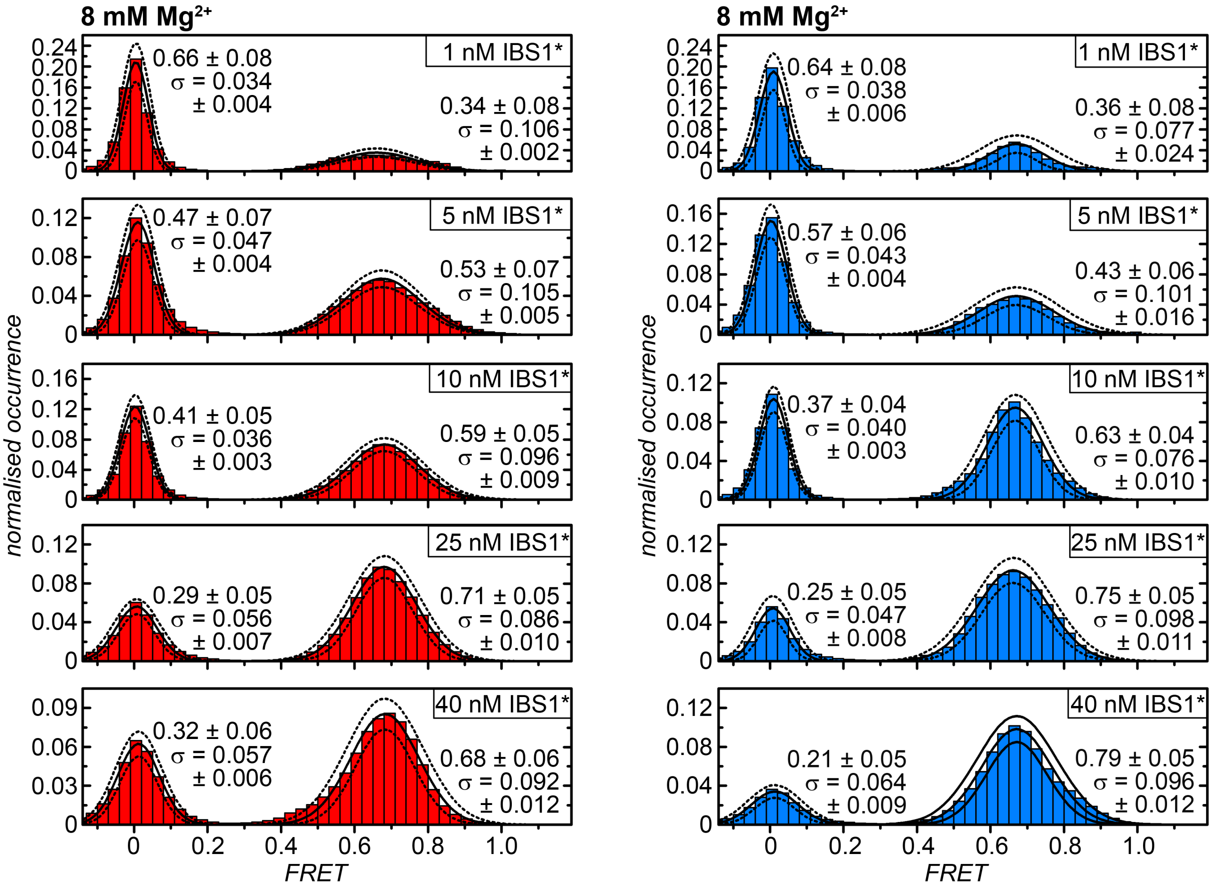


Figure 5.3: Normalised averaged histograms built from smFRET time traces recorded at constant $M^{2+}$ concentration and increasing IBS1* concentration. Increasing IBS1* concentration favors the occurrence of the d3'EBS1*/IBS1* complex, *i.e.* the *high FRET* state. Values for bound and unbound fractions, and errors (2$\sigma$ determined by bootstrapping and regression using BOBA FRET (13). Imaging conditions: 50 mM MOPS, 100 mM KCl, 8 mM $MgCl_2$, 1 % (w/V) D-Glucose, 1x OSS, 1 mM Trolox, pH 6.90, IBS1* as indicated.

### 5.2.9 Kinetic rates

Cumulative probability distributions $P(t_{\text{zero}})$ and $P(t_{\text{high}})$ were built according a method described by Hellriegel *et al.* [395]. $P$ was always normalised to 1, $t_{\text{zero}}$ and $t_{\text{high}}$ refer to the dwell-times of molecules in the undocked and the docked *FRET* state with $t_{\text{zero}} = t_{\text{zero},i,n}$ and $t_{\text{high}} = t_{high,j,n}$ for all $i, j$ and $n$ recorded under equal buffer conditions. Complementary cumulative probability distributions $C$ were determined according to:

$$C(t_{\text{zero}}) = 1 - P(t_{\text{zero}}) \tag{5.13}$$

and

$$C(t_{\text{high}}) = 1 - P(t_{\text{high}}) \tag{5.14}$$

Subsequently, these distributions were fitted to exponential decay functions [204, 195, 219]:

$$C(t_{\text{zero}}) = \sum_{m=1}^{m_{\max}} a_{\text{zero},m} \exp\left\{-\frac{t_{\text{zero}}}{\tau_{\text{zero},m}}\right\} \tag{5.15}$$

$$C(t_{\text{high}}) = \sum_{m=1}^{m_{\max}} a_{\text{high},m} \exp\left\{-\frac{t_{\text{high}}}{\tau_{\text{high},m}}\right\} \tag{5.16}$$

where $m_{\max}$ denotes the total number of exponential decay functions required to satisfactorily describe the experimental data. $a_{\text{zero},m}$ and $a_{\text{high},m}$ refer to the corresponding amplitudes. The kinetic rate constants $k_{\text{dock}}$ and $k_{\text{undock}}$ were determined from the resulting time constants $\tau_{\text{zero}}$ and $\tau_{\text{high}}$:

$$k_{\text{undock},m} = \frac{1}{\tau_{\text{high},m}} \tag{5.17}$$

$$k_{\text{dock},m} = \frac{1}{\tau_{\text{zero},m} c(\text{IBS1}^*)} \tag{5.18}$$

Affinity/dissociation constants in the absence of $M^{2+}$ were then calculated according to

$$K_{\text{A},m} = \frac{k_{\text{dock},m}}{k_{\text{undock},m}} = \frac{\tau_{\text{high},m}}{\tau_{\text{zero},m} c(\text{IBS1}^*)} = \frac{1}{K_{\text{D},m}} \tag{5.19}$$

### 5.2.10 Calculation of $K_{\text{A},n}$ distributions

To predict the fraction of d3'EBS1* molecules in the undocked state $P(<0.375)$ at variable $c(\text{IBS1*})$, the heterogeneity of the binding constants is accounted for by introducing $l_{total}$ subspecies with binding constants $K_{\text{A},l}$ that occur in certain fractions $f_l$. These conditions are implemented by transposition and modification of Equation 5.4:

$$\frac{p(FRET_{\text{zero}})}{p(FRET_{\text{zero}})+p(FRET_{\text{high}})}=\frac{t_{\text{zero}}}{t_{\text{zero}}+t_{\text{high}}}=\sum_{i=l}^{l_{total}}\frac{f_l}{K_{\text{A},l}c(\text{IBS1}^*)+1} \tag{5.20}$$

with $\sum f_l = 1$. Please refer to the next section for a description of the approach employed to extract discrete $K_{\text{A},l}$ from statistically broadened cumulative probability distributions $\text{P}(K_{\text{A},n})$. It must be emphasised that the obtained values $K_{\text{A},l}$ may not be regarded as discrete. Rather, they are average values over a more or less continuous distribution of molecules, where $f_l$ refers to the fraction of molecules close to the value $K_{\text{A},l}$ [387]. The averaged undocked fractions of d3'EBS1* for several hundreds of molecules ($n_{\text{max}} > 200$) were determined with $t_{\text{high}} = t_{\text{high}}^{\text{sum}}$ and $t_{\text{zero}} = t_{\text{zero}}^{\text{sum}}$ (see Equations 5.10 and 5.11). Equation 5.20 holds true when (i) d3'EBS1* immobillisation does not interfere with IBS1* docking/undocking, (ii) the underlying mechanism can be quantitatively described by a standard bimolecular reaction, and (iii) $c(\text{IBS1*})_{\text{total}} >> c(\text{d3'EBS1*})$. The docked fraction is calculated in an analogous manner:

$$\frac{p(FRET_{\text{high}}}{p(FRET_{\text{zero}})+p(FRET_{\text{high}})}=\frac{t_{\text{high}}}{t_{\text{zero}}+t_{\text{high}}}=\sum_{i=l}^{l_{\text{total}}}\frac{f_l}{K_{\text{A},l}c(\text{IBS1}^*)+1} \tag{5.21}$$

### 5.2.11 Simulation of dwell times

Simulations were performed in Matlab (release 2013b, license of the University of Zurich) using the built-in function "random('exp',*m*,*n*)", which creates an *n*-by-*m* matrix of exponentially distributed numbers.

### 5.2.12 Nuclear magnetic resonance (NMR)

All spectra were recorded on a Bruker Avance 700 MHz spectrometer with a 5 mm CRYO TXI inverse triple-resonance probehead with *z*-gradient coil. $\text{Mg}^{2+}$-induced chemical shift changes $\Delta\delta$ were determined by adding 0, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 6.5 and 7 mM $\text{MgCl}_2$ to d3'EBS1* or d3'EBS1*/IBS1* dissolved in $\text{D}_2\text{O}$, respectively. Upon each titration step, a $[^1\text{H},^1\text{H}]$-NOESY spectrum (d3'EBS1*: 293 K, 10 mM KCl, 10 µM EDTA; d3'EBS1*/IBS1*: 298 K, 110 mM KCl, 10 µM EDTA) was recorded as described in [50]. The $\Delta\delta$ values corresponding to ribose H1' protons and base H5, H6, H8 and H2 protons were fitted to a 1:1 binding isotherm for each $\text{Mg}^{2+}$ concentration:

$$\begin{aligned}\delta_{obs} =&\delta_{\text{RNA}}+(\delta_{\text{RNA-M}}^{2+}-\delta_{\text{RNA}})(\frac{c(\text{M}^{2+}+c(\text{RNA})+K_{\text{D}}}{2c(\text{RNA})}\\ &-\frac{\sqrt{(c(\text{M}^{2+})+c(\text{RNA})K_{\text{D}})^2-4c(\text{M}^{2+}c(\text{RNA})}}{2c(\text{RNA}))}\end{aligned} \tag{5.22}$$

where $\delta_{\text{RNA}}$ is the chemical shift value of the proton in the absence of divalent metal ions and

$\delta_{\text{RNA-Mg2+}}$ is the chemical shift of the proton in the RNA-metal complex [76]. Poorly fitted titration curves were discarded and protons with similar $K_A$ values were grouped according to the previously defined $Mg^{2+}$-binding sites [50]:

**d3'EBS1***
**Stem:** A3H1', G4H1', G4H8, U5H1', U5H6, A25H8, C26H1', C26H6, U27H1', U27H6
**Loop:** A10H1', A10H2, A10H8, U11H1', U1H5, U11H6, U12H1', U12H6, G21H1', A20H1', A20H2, A20H8

**d3'EBS1*/IBS1***
**Stem:** A3H1', G4H1', G4H8, U5H6, A25H2, C26H1', C26H6, U27H6
**Loop 1:** G13H1', G13H8, G14H1', G14H8
**Loop 2:** A16H2, A16H8, C17H6, U18H6, G19H1', A20H1', G21H1', G21H8, C59H1', C59H5, C59H6, A60H2, A60H8, G61H1', G61H8, U62H6, G63H1', G63H8, U64H1'

The average association constants of each binding site were refined iteratively using the Matlab script ISTAR v2.3 as described in [48]. $Mn^{2+}$-induced line broadening was followed in a series of [$^1$H,$^1$H]-NOESY spectra recorded in 100 % $D_2O$ at 293 K for d3'EBS1* and 298 K for d3'EBS1*/IBS1*. $MnCl_2$ was added in steps of 0, 10, 30, 60, 90, 120, 150, 200 and 300 μM (d3'EBS1*), and 0, 30, 60, 90, 120, 180, 240, 300 and 360 μM (d3'EBS1*/IBS1*). Line broadening of individual peaks was assessed visually using Sparky (http://www.cgl.ucsf.edu/home/sparky/). NOE crosspeaks between $[Co(NH_3)_6]^{3+}$ and d3'EBS1* or d3'EBS1*/IBS1* were assigned in [1H,1H]-NOESY spectra recorded in 90 % $H_2O$/10 % $D_2O$ at 278 K.

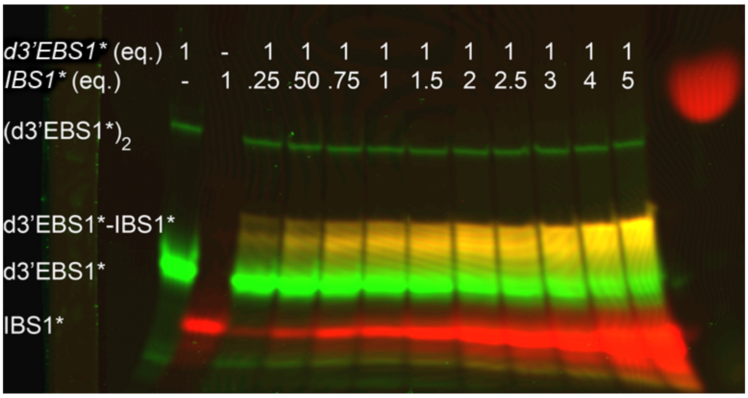


Figure 5.4: d3'EBS1*/IBS1* affinity in the presence of $Mg^{2+}$ and increasing IBS1* concentration as assessed with native gel electrophoresis, representative data (overlay). Four conformational species are observed and assigned as follows: $(d3'EBS1^*)_2$ (upper green band), d3'EBS1*-IBS1* (yellow band), undocked d3'EBS1* (lower green band) and free IBS1* (red band). Migration performed at 5 mM MOPS, 2 mM NaOH, 50 μM $MgCl_2$, pH 6.90.

## 5.3 Results

### 5.3.1 Divalent metal ions enhance d3'EBS1*/IBS1* affinity

The d3'EBS1*/IBS1* interaction is strictly dependent on $Mg^{2+}$ concentration [388]. Native PAGE with the Cy3-d3'EBS1*/Cy5-IBS1* system shows that (i) the attached fluorophores do not hamper strand association, and (ii) that $M^{2+}$ ions further promote their interaction (Figure 5.4). For an in-depth characterisation, single molecule fluorescence traces over several minutes were recorded for the Cy3-d3'EBS1*/Cy5-IBS1* docking/undocking, varying either the $M^{2+}$ or the IBS1* concentration.

The time traces reveal two well-separated *FRET* states: a zero *FRET* state (undocked d3'EBS1*) and a high *FRET* (docked d3'EBS1*/IBS1*) (Figure 5.1B). Control experiments confirmed 1:1 binding stoichiometry and the negligible influence of bleaching and long-lived dark states under the imaging conditions chosen herein (Figs. 5.5 and 12.10 in the Appendix). Also, tethering fluorophores to RNA and the presence of $M^{2+}$ ions have little to no effect on their fluorescence spectra, quantum yields and rotational diffusion (Figures 12.4, 12.5, 12.6, 12.7, 12.8 and 12.9 in the Appendix). Trajectories were classified as dynamic (type I and II) or static (type IIIa and IIIb) depending on whether interconversion events between these states were present or absent during the observation time of 400 seconds. Static undocked molecules (type IIIa) were excluded from cumulated normalised *FRET* histograms (Figures 5.2 and 5.3).

In the absence of $M^{2+}$, the high *FRET* peak centers at $\sim$ 0.6 and is populated during only 5 % of the accumulated observation time (Figure 5.6A). Moreover, $>$ 60 % of all molecules

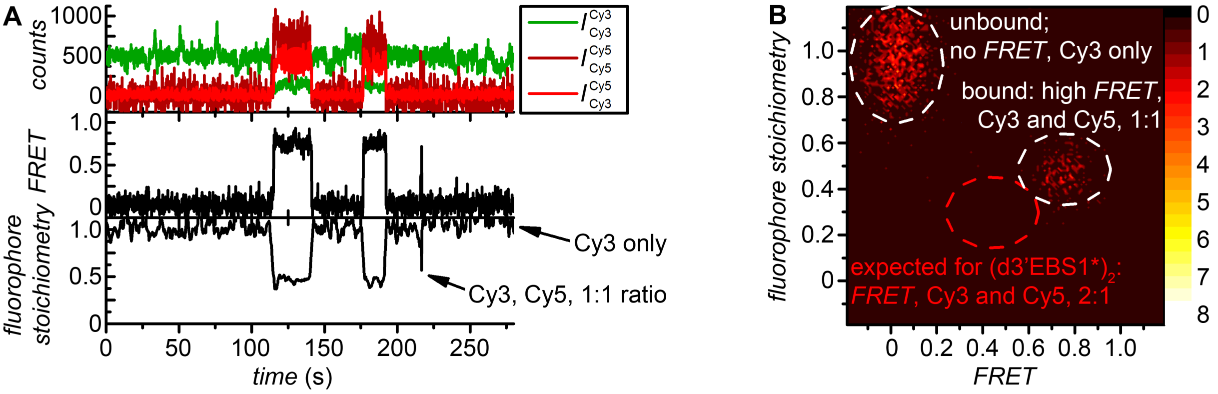


Figure 5.5: smFRET control experiments. (A) Top: ALEX yields three photon counts over time. Middle: *FRET* over time calculated according to Equation 5.1. Bottom: Fluorophore *stoichiometry* as determined according to Equation 5.2 varies between 1 (undocked state), and 0.5 (docked state). Trace smoothed with sliding average over 10 data points. (B) 2D plot built from the data shown in (A). As expected from the experimental design, two populations are observed [396]. IBS1* docked onto d3'EBS1* dimers (red area) or intermediate FRET efficiencies are not observed. The colour code represents the number of counts. Imaging conditions: 50 mM MOPS, 100 mM KCl, 25 nM IBS1*, 1 % (w/V) D-Glucose, 1x OSS, 1 mM Trolox, pH 6.90.

remain statically undocked (type IIIa, Figure 5.7). Upon addition of either 1 mM $Mg^{2+}$ or $Ca^{2+}$, the high *FRET* state becomes significantly more populated with a roughly 1:1 ratio of docked and undocked state (Figs. 5.2 and 5.6B) and the fraction of type IIIa molecules is reduced to ~ 20 % (Figure 5.7). At the same time, the half width of the high *FRET* state is considerably narrowed (Figures 5.6D and 5.2). Addition of 2-8 mM $M^{2+}$ does not further decrease the fraction of type IIIa molecules (Figure 5.7), though, it leads to a higher occurrence of the high-FRET state, an effect that is more pronounced for $Ca^{2+}$ than $Mg^{2+}$ (Figure 5.6C). Simultaneously, the maxima of each high *FRET* state shift towards higher values, *i.e.* from 0.60 to 0.75 in the case of $Mg^{2+}$ and reaching 0.70 with $Ca^{2+}$ (Figure 5.6D). The higher propensity of $Ca^{2+}$ to stabilise the interaction of the d3'EBS1*/IBS* complex was confirmed in further smFRET experiments conducted at constant $M^{2+}$ concentration: At 8 mM $Mg^{2+}$, and from 1-40 nM IBS1*, the fraction of the docked state increases to ~ 70 % whereas almost 80 % of all d3'EBS1* molecules are docked at 8 mM $Ca^{2+}$ and 40 nM IBS1* (Figure 5.3).

The single molecule data clearly show that 100 mM $K^+$ alone are not sufficient for d3' EBS1*/IBS1* recognition and stable docking (Figure 5.6), and that the fraction of type IIIa molecules is high. In turn, addition of physiological amounts of $M^{2+}$ leads to stable interaction and a strong decrease in the fraction of static undocked molecules. Whether these inactive molecules display a much lower $K_{\mathrm{A}}$ or they correspond to a photophysical artifact was not assessed fur-

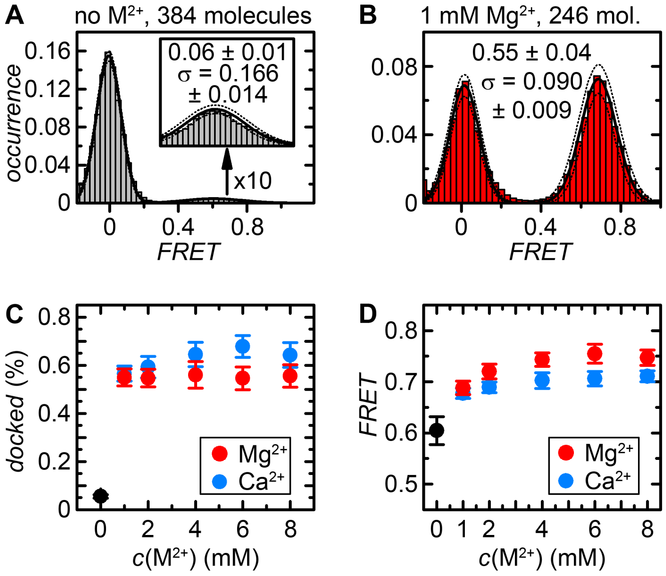


Figure 5.6: Analysis of cumulated normalised *FRET* histograms. (A-B) Representative data at 0 mM and 1 mM $Mg^{2+}$. The inset in (A) shows a 10-fold magnification of the high FRET distribution. Solid lines correspond to Gaussian fits, dotted lines indicate twice the bootstrapped standard deviation of the amplitude and the width as determined using the Matlab script BOBA FRET (number of replicates = number of molecules, number of bootstrap samples = 100, for details see ref. [166]). (C) Normalised occurrence of docked d3'EBS1* at different $M^{2+}$ concentration as determined through thresholding at *FRET* = 0.375. (D) High *FRET* distribution maxima in the averaged histograms. All errors correspond to $2\sigma_{bootstrap}$ [166].

ther, albeit 15-55 % of the total population is routinely found to be static in smFRET studies using the FRET pair Cy3 and Cy5 [146]. Addition of only 1 mM $M^{2+}$ also narrows the distribution of the high *FRET* state considerably (Figure 5.6B). Since d3'EBS1*/IBS1* has recently been shown to adopt a rigid structure even in the absence of $M^{2+}$, it is unlikely that the broad *FRET* distribution is due to (hidden) molecular motion [50, 146]. Rather, docking events are thought to be shorter than the experimental time resolution, leading to a decrease in the high *FRET* value. In turn, the slight decrease in mean interdye distance with increasing $M^{2+}$ concentrations reflects most likely a structural compaction of the d3'EBS1*/IBS1* complex, which is consistent with increased charge screening [230].

### 5.3.2 Heterogeneity in d3'EBS1*/IBS1* interaction kinetic is increased by divalent metal ions

In order to gain insights into the kinetics of docking and undocking, we analysed the distribution of dwell times. As a dwell time must have clearly defined start and end points, only time traces with at least two transitions were included in the analysis (type I). Generally, we find molecules with rapid docking/undocking behavior as well as molecules that stay for many seconds or even minutes in either state (Figure 5.1B). There is no classification into distinct kinetic subspecies possible [213], though, strand docking/undocking kinetics are found to be generally accelerated at increasing $Mg^{2+}$ or $Ca^{2+}$ concentration (Figure 5.8).

In the absence of $M^{2+}$, a mono-exponential decay model satisfactorily describes the normalised complementary cumulative probabilities of dwell times in the undocked state (Figure 5.9A and Equation 5.15, $m_{max}$ = 1).The average dwelltime in the undocked state $\tau_{\text{zero}}$ is 76 s, corresponding to a docking rate $k_{\text{dock}} = 0.52\ \mu M^{-1}\ s^{-1}$ (Equation 5.18). In contrast, the cumulative probabilities of dwell times in the docked state can only be fitted with a multiexponential model (Figure 5.9B and Equation 5.16, $m_{\text{max}} = 3$). The average time spent in the docked state $\tau_{\text{high}}$ is 1.2 s (22 %), 8.2 s (70 %), and 30 s (8 %), respectively. Based on the docking and undocking rates, equilibrium constants ($K_A$, $K_D$) and Gibbs free energy values ($\Delta G°$) were calculated for the three fractions using Equation 5.19 (Table 5.2). The averaged $\log K_A = 6.6$ ($\Delta G° = 37$ kJ

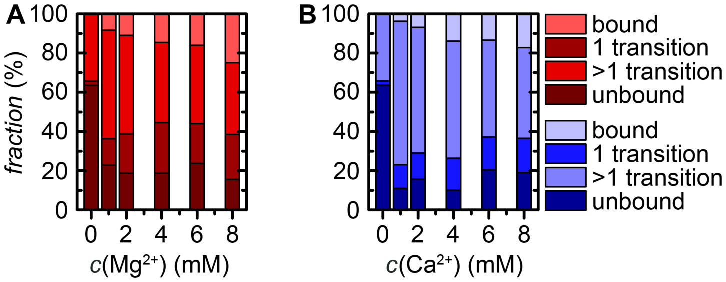


Figure 5.7: Fractions of static and dynamic molecules in the presence of $Mg^{2+}$ or $Ca^{2+}$. Time traces were classified as type I (>1 transition), type II (1 transition), type IIIa (static undocked), and type IIIb (static docked).

mol$^{-1}$) and is in perfect agreement with the corresponding bulk value as obtained from thermal melting studies ($\Delta G^{\circ}$ = 38.7 0.9 kJ mol$^{-1}$), demonstrating that surface immobilisation does not adversely affect d3'EBS1*/IBS1* interaction (Table 5.3) [200]. Upon addition of Mg$^{2+}$, the fraction of type IIIa molecules is consistently reduced to 20 % at the concentration range applied (Figure 5.7A). The presence of Ca$^{2+}$ similarly decreases the number of static molecules, although the number of dynamic molecules (types I and II) generally seems to be higher than in the presence of Mg$^{2+}$ (Figure 5.7). Simultaneously, both the addition of 1 mM Mg$^{2+}$ and Ca$^{2+}$ leads to a decrease in $\tau_{\mathrm{zero}}$, whereas $\tau_{\mathrm{high}}$ strongly increases, a trend that persists at higher M$^{2+}$ concentration (Figure 5.9). These data demonstrate that M$^{2+}$ ions promote IBS1* docking by accelerating association and slowing down dissociation, leading to a shift of the thermodynamic equilibrium towards the docked state, consistent with the FRET histogram data.

Multiexponential decay functions were required throughout to obtain satisfactory fits to the experimental data (Equations 5.15 and 5.16, $m_{\mathrm{max}}$ = 3, Table 5.4, Figure 5.9). Binding constants could not be calculated based on the ratio of $k_{\mathrm{dock}}$ and $k_{\mathrm{undock}}$, because (i) the individual rate constants for docking and undocking were not constant at different M$^{2+}$ concentrations, (ii) no systematic trend of rate constants and/or amplitudes was observed (Table S4), (iii) various combinations of $\langle t_{\mathrm{zero},n} \rangle$ and $\langle t_{\mathrm{high},n} \rangle$ are observed and the corresponding $K_{\mathrm{A},n}$ values span (Equation 5.7) a continuum of more than three orders of magnitude (Figure 5.10A,B). While such scatter plots clearly illustrate which combinations of docking and undocking rates oc-

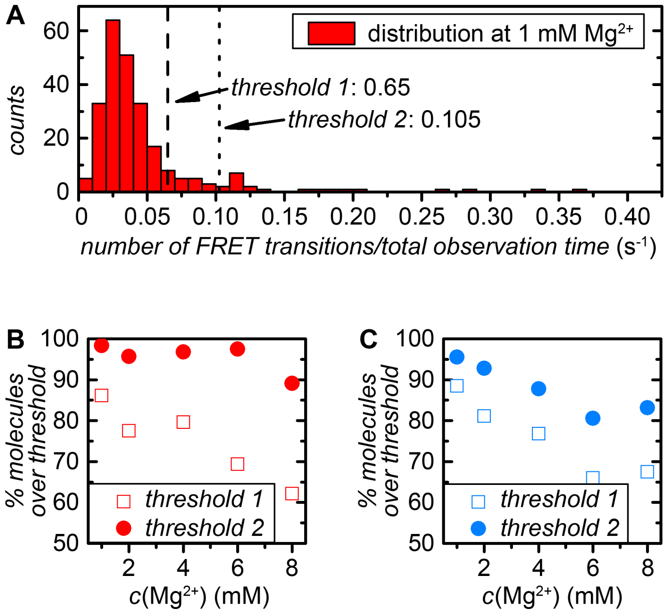


Figure 5.8: d3'EBS1*/IBS1* docking and undocking rates at different divalent metal ion concentrations. (A) Histogram showing interconversion rates calculated as the number of *FRET* transitions over the total observation time. Two arbitrary thresholds to distinguish rapidly and slowly interconverting molecules were defined (dashed and dotted line). (B,C) Fraction of slowly interconverting molecules at different Mg$^{2+}$ and Ca$^{2+}$ concentrations.

Table 5.2: Dwell-time analysis of d3'EBS1*/IBS1* docking/undocking in the absence of divalent metal ions. $\log K_\text{A}$ and $K_\text{D}$ were calculated according to Equation 5b with $\langle t_\text{high} \rangle = \tau_{\text{high},m}$ and $\langle t_\text{zero} \rangle = \tau_\text{zero}$, whereas $\tau_\text{high}$ and $\tau_\text{zero}$ result from mono- and triexponential fits according to Equations 5.15 and 5.16 of the complementary cumulative probability distribution of dwell-times shown in Figure 5.9. Only type I molecules were taken into account.

| $a_{\text{high},m}$ (%) | $\tau_{\text{high},m}$ (s) | $\tau_\text{zero}$ (s) | $\log K_{\text{A},m}$ | $K_{\text{D},m}$ (nM) | $\Delta G^\circ_m$ (kJ mol$^{-1}$) | $k_\text{dock}$ (μM$^{-1}$ s$^{-1}$) | $k_\text{undock}$ (s$^{-1}$) |
|---|---|---|---|---|---|---|---|
| 22 | 1.2 | 76.5 | 5.8 | 1580 | -33.0 | 0.52 | 0.83 |
| 70 | 8.2 | 76.5 | 6.6 | 240 | -37.8 | 0.52 | 0.12 |
| 8 | 30.0 | 76.5 | 7.2 | 66 | -41.1 | 0.52 | 0.03 |

cur, the distribution of $K_{\text{A},n}$ is better described by cumulative probability functions $P(\log K_{\text{A},n})$ revealing a broad sigmoid covering several orders of magnitude in this semi-logarithmic plot (Figure 5.10C,D).

Table 5.3: Thermodynamic parameters of the d3'EBS1*/IBS1* interaction in the presence of divalent metal ions as determined by UV thermal melting and smFRET. Errors correspond to the standard deviation ($1\sigma$).

| $c(\text{M}^{2+})$ | $\Delta\text{G}^\circ_\text{UV}$ (kJ mol$^{-1}$) | $\Delta\text{G}^\circ_\text{smFRET}$ (kJ mol$^{-1}$) | $\log K_\text{A, smFRET}$ | $K_\text{D, smFRET}$ (nM) |
|---|---|---|---|---|
| 0 | -38.7 ± 0.9 | -37.9 | 6.6 | 225 |
| 1 mM $\text{Mg}^{2+}$ | -41.6 ± 1.4 | -45.6 | 8.0 | 10.1 |
| 1 mM $\text{Ca}^{2+}$ | -41.1 ± 1.0 | -44.0 | 7.7 | 19.6 |
| 2 mM $\text{Mg}^{2+}$ | -45.4 ± 0.5 | -45.2 | 7.9 | 11.8 |
| 2 mM $\text{Ca}^{2+}$ | -43.6 ± 1.4 | 44.4 | 7.8 | 16.7 |
| 4 mM $\text{Mg}^{2+}$ | -46.4 | -46.4 | 8.1 | 7.4 |
| 4 mM $\text{Ca}^{2+}$ | -45.5 ± 0.3 | -46.5 | 8.2 | 6.9 |
| 6 mM $\text{Mg}^{2+}$ | -47.3 | -45.6 | 8.0 | 10.1 |
| 6 mM $\text{Ca}^{2+}$ | -46.3 ± 1.7 | -45.8 | 8.0 | 9.4 |
| 8 mM $\text{Mg}^{2+}$ | -48.9 ± 0.4 | -46.6 | 8.2 | 6.7 |
| 8 mM $\text{Ca}^{2+}$ | -48.4 ± 0.3 | -46.3 | 8.1 | 7.6 |

### 5.3.3 Approximation of $K_A$ distributions broadened by kinetic heterogeneity

The dwell-times $t_{\mathrm{zero},n,i}$ and $t_{\mathrm{high},n,j}$ are exponentially distributed according to the statistical nature of binding and folding processes [195, 219, 397]. Therefore, histograms of $t_{1,n,i}$ and $t_{2,n,j}$ are typically fitted to exponential models [195, 219]. In heterogeneous systems, the dwell-time histograms require more than one exponential fit functions to be approximated [396]. As $K_{\mathrm{A},n}$ depends directly on the average dwell times $\langle t_{\mathrm{zero},n} \rangle$ and $\langle t_{\mathrm{high},n} \rangle$ in the undocked and docked state (Equation 5.21), the uncertainty of $K_{\mathrm{A},n}$ depends on $i_{\max}$ and $j_{\max}$ (29). Finite length of smFRET time traces is therefore expected to broaden the $K_{\mathrm{A},n}$ frequency distribution, even for a homogeneous system with a single $K_A$ [224].

#### A mathematical description of the width of $K_{\mathrm{A},n}$ distributions for a homogeneous system and its dependence on $i_{\max}$ and $j_{\max}$

Cumulative probability distributions $P(K_{\mathrm{A},n})$ are binning-free and make use of the same formalism as for the dwell-time distributions $P(t_{\mathrm{zero/high}})$ [395]. According to Equation 5.21, $K_{\mathrm{A},n}$ depends on three parameters: $c$(IBS1*) (constant), whereas $\langle t_{\mathrm{zero},n} \rangle$ and $\langle t_{\mathrm{high},n} \rangle$ display inherent statistical fluctuations. Given the assumption that the system is homogeneous and the principle of ergodicity applies [161], it is possible to determine the overall $K_A$ by fitting the

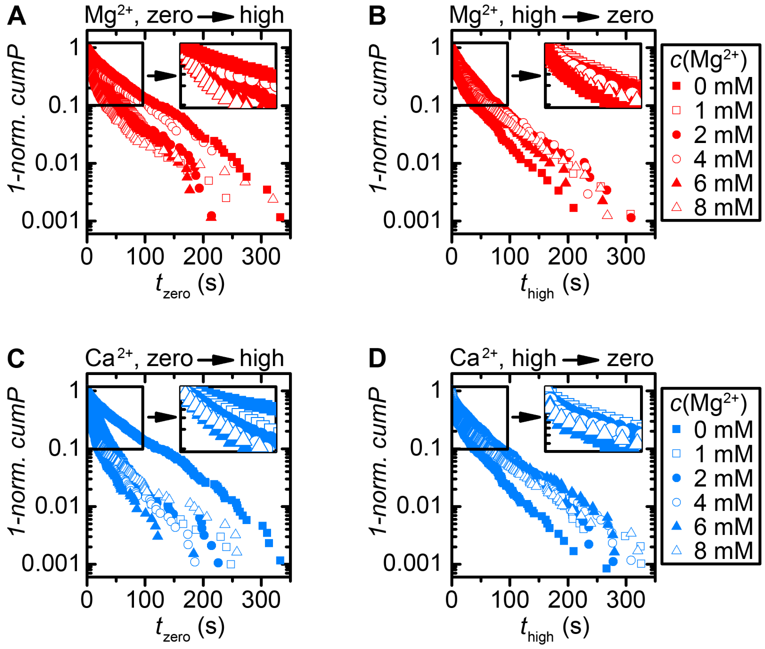


Figure 5.9: Cumulative probability distributions of dwell times showing the decay of the zero FRET state ($C(t_{\mathrm{zero}})$, see Equation 8a) and the decay of the high FRET state ($C(t_{\mathrm{high}})$, see Equation 8b) in the presence of $Mg^{2+}$ or $Ca^{2+}$. The inset shows a magnification for short dwell times. $C(t_{\mathrm{zero}})$ in the absence of divalent metal ions can be fitted to a monoexponential decay function, all other experimental data sets require multiexponential fits. 50 % of all data points were omitted for clarity.

distribution $p(K_{A,n})$ or $P(K_{A,n})$ from a sufficiently high number individual molecules with an appropriate analytical function.

### Case 1: $i_{\max} = j_{\max} = 1$

Here, the distribution of $p(K_{A,n})$ corresponds to the ratio distribution of two exponential distributions $p(\langle t_{\text{high},j,n}\rangle)$ and $p(\langle t_{\text{zero},i,n}\rangle)$ with the analytical solution:

$$p(K_{\text{A},n}) = \frac{p(\langle t_{high,n}\rangle_{j=1})}{[\text{IBS1}^*]p(\langle t_{\text{zero},n}\rangle_{i=1})} = \frac{[exp]_{\tau_{\text{high}}}}{[\text{IBS1}^*[exp]_{\tau_{\text{zero}}}} = A\frac{R}{(1+K_{\text{A},n}R)^2} \tag{5.23}$$

In general, the time constants $\tau_{\text{high}}$ and $\tau_{\text{zero}}$ are the expected values of dwell times in state 1 and state 2, $A$ is an amplitude to account for the number of events in an experimental and unnormalised frequency distribution and the true $K_A$ can be derived by $K_A = 1/R$. Please refer to Figure 5.11A (top) for a visual representation of this case. Please note that $R \sim \tau_{\text{zero}}/\tau_{\text{high}}$.

Table 5.4: Dwell-time analysis of d3'EBS1*-IBS1* docking/undocking in the presence of divalent metal ions using triexponential fits. $\tau_{high,m}$ and $\tau_{zero,m}$ ($m_{max}$ = 3, see Equations 5.15 and 5.16) result from the exponential fits to the complementary cumulative probability distributions of dwell-times shown in Figure 5.9.

| | fast decay (s) | | intermediate decay (s) | | slow decay (s) | |
|---|---|---|---|---|---|---|
| $c(\text{M}^{2+})$ | $\tau_{\text{high},1}$ | $\tau_{\text{zero},1}$ | $\tau_{\text{high},2}$ | $\tau_{\text{zero},2}$ | $\tau_{\text{high},3}$ | $\tau_{\text{zero},3}$ |
| 0 | 1.2 | 76.5 | 8.2 | - | 30.0 | - |
| 1 mM $\text{Mg}^{2+}$ | 0.8 | 2.0 | 18.5 | 22.7 | 48.3 | 42.0 |
| 1 mM $\text{Ca}^{2+}$ | 1.0 | 3.5 | 1.9 | 16.9 | 16.9 | 65.7 |
| 2 mM $\text{Mg}^{2+}$ | 1.8 | 5.2 | 9.9 | 24.2 | 46.9 | 126.1 |
| 2 mM $\text{Ca}^{2+}$ | 10.4 | 1.1 | 42.3 | 10.3 | - | 54.7 |
| 4 mM $\text{Mg}^{2+}$ | 6.5 | 0.5 | 21.7 | 6.9 | 88.5 | 60.4 |
| 4 mM $\text{Ca}^{2+}$ | 2.6 | 1.2 | 11.0 | 3.2 | 55.2 | 52.1 |
| 6 mM $\text{Mg}^{2+}$ | 0.7 | 2.7 | 7.9 | 11.3 | 50.5 | 58.4 |
| 6 mM $\text{Ca}^{2+}$ | 1.5 | 1.7 | 6.5 | 14.6 | 34.6 | 77.9 |
| 8 mM $\text{Mg}^{2+}$ | 0.8 | 0.8 | 7.3 | 7.3 | 55.4 | 55.4 |
| 8 mM $\text{Ca}^{2+}$ | 1.6 | 0.6 | 8.7 | 4.9 | 67.6 | 44.5 |

**Case 2: $i_{\max} = j_{\max} > 1$**

Here, $p(\langle t_{\mathrm{high},j,n}\rangle)$ and $p(\langle t_{\mathrm{zero},i,n}\rangle)$ are no longer exponentially distributed, since the exponentially distributed values are averaged over $i_{\max}$ or $j_{\max}$ number of dwell times. Again, $\langle t_{\mathrm{zero,i_{max},}n}\rangle$ corresponds to the average time a molecule $n$ dwells in a certain *FRET* state (Equations 5.11 and 5.12). Thus the distribution of $p(K_{\mathrm{A},n})$ is no longer the ratio of two exponential distributions, but the ratio distribution of two gamma distributions with expected values $\tau_{\mathrm{high}}$ and $\tau_{\mathrm{zero}}$:

$$p(K_{\mathrm{A},n}) = \frac{p(\langle t_{\mathrm{high},n}\rangle_{i\geq 1})}{[\mathrm{IBS1}^*]p(\langle t_{\mathrm{zero},n}\rangle_{i\geq 1})} = \frac{[gamma]_{\tau_{\mathrm{high}}}}{[\mathrm{IBS1}^*[gamma]_{\tau_{\mathrm{zero}}}} \tag{5.24}$$

Averaging monoexponentially distributed dwell times yields the probability distribution $p(\langle t_{\mathrm{zero},\, i_{\max,n}}\rangle)$, which is analytically described by a gamma distribution [396, 398]. In this case, it is defined as follows:

$$p_{\Gamma}(\langle t_{\mathrm{zero},i_{\max,n}}\rangle) = A\frac{k^{i_{\max}}\langle t_{\mathrm{zero},i_{max,n}}\rangle^{i_{\max}-1}}{\Gamma(i_{\max})}\exp\left\{-k\langle t_{\mathrm{zero,i_{max,}}n}\rangle\right\} \tag{5.25}$$

where $k = 1/\tau_{\mathrm{zero}}$, $A$ is an amplitude, $\Gamma$ is the gamma function, and $i_{\max}$ the number of dwell

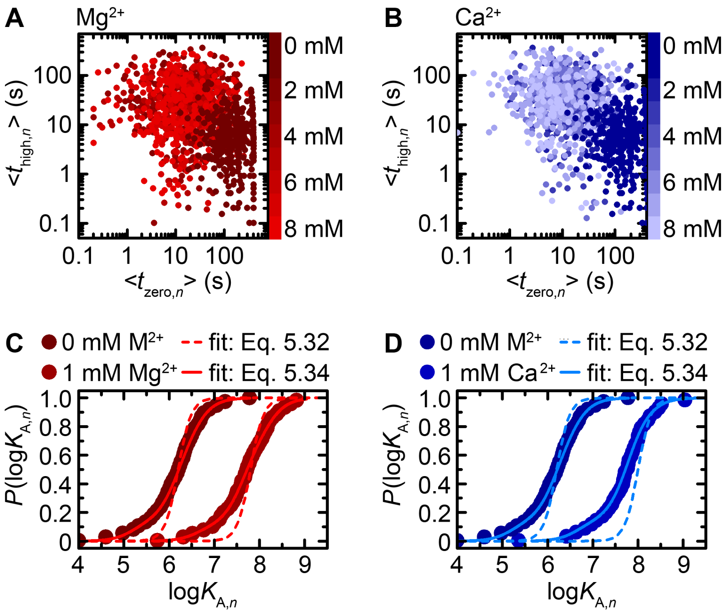


Figure 5.10: Distribution of single-molecule $K_{\mathrm{A}}$ values in the presence of 1-8 mM $\mathrm{Mg}^{2+}$ and $\mathrm{Ca}^{2+}$ as determined by Equation 5b. (A-B) Scatter plots of the average dwell time in the docked state against the average time spent in the undocked state. The colour code represents the $\mathrm{M}^{2+}$ concentration. (C-D) Normalised cumulative probability plots of $\log K_{\mathrm{A},n}$ values in the presence of 0 and 1 mM $\mathrm{Mg}^{2+}$ or $\mathrm{Ca}^{2+}$. Dotted lines assume homogeneous interaction kinetics ($i_{\max} = 4$, $j_{\max} = 4$, Equation 15c). Solid lines represent fits to the experimental data assuming three subpopulations ($\log K_{\mathrm{A,av}} = 6.29$, 7.85, 7.78, respectively, $s_{\max} = 3$, Equation 17b, see also Table S4).

times used for averaging, which ultimately defines the width of the Gamma distribution. In order to quantitatively assess case 2, we computed dwell-time histograms $p(\langle t_{\mathrm{dwell},i_{\max,n}}\rangle)$ from 10000 simulated smFRET time traces with $i_{\max} = j_{\max} = 1, 3, 5, 10, 30$ or 90 and fitted them to Equation 5.25. Figure 5.11A clearly shows that Equations 5.23 and 5.25 describe the simulated distributions $p(\langle t_{\mathrm{zero},i_{\max,n}}\rangle)$ very well. Please note that high values for $i_{\max}$ and $j_{\max}$

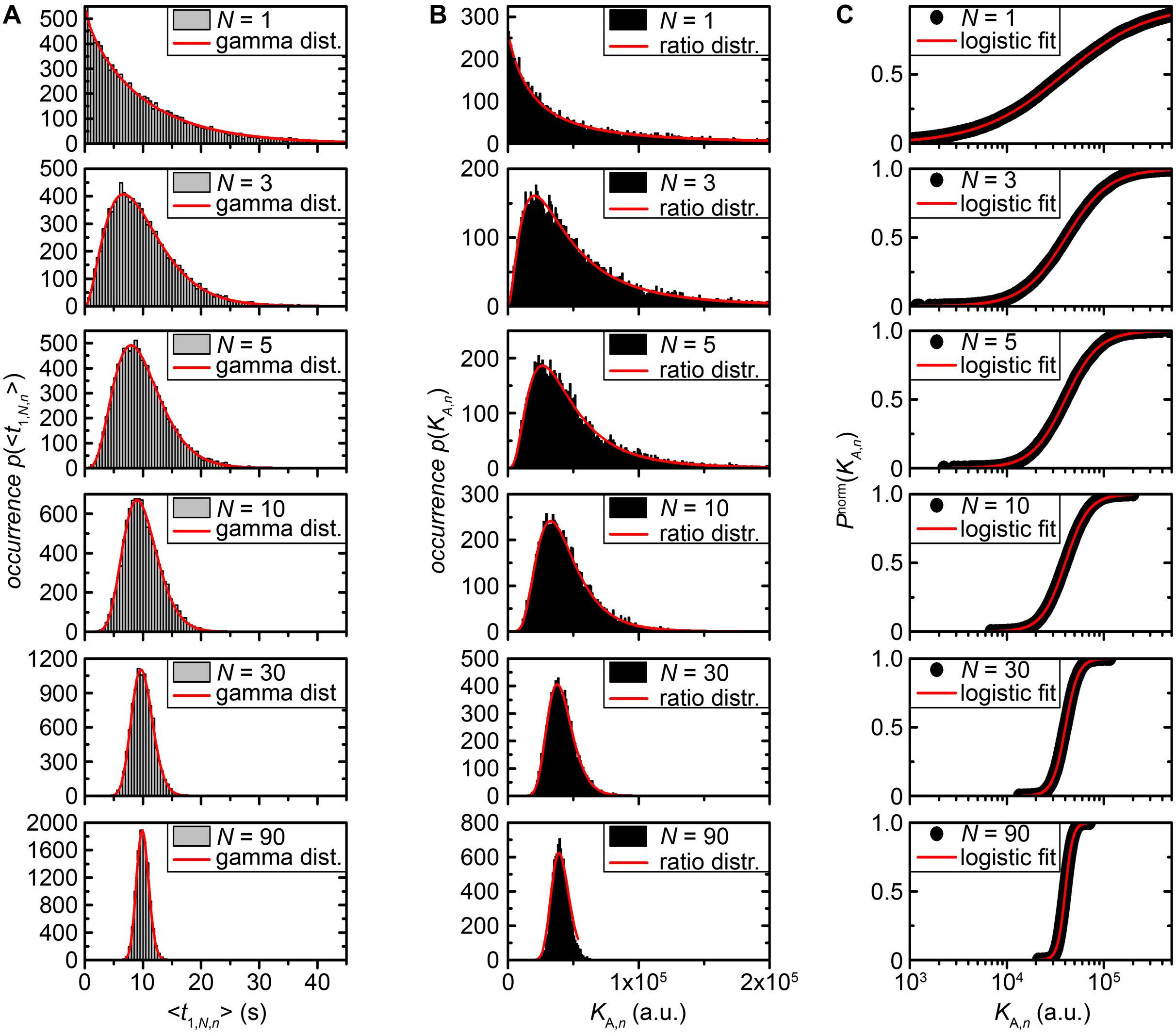


Figure 5.11: Simulations of dwell-time and $K_\mathrm{A}$ distributions. (A) Dwell time histograms built from 10000 simulated single-molecule FRET time traces with $N := i_{\max} = j_{\max} = 1, 3, 5, 10, 30$, and 90 exponentially distributed dwell times. The red curves correspond to a gamma distribution fit using Equation 5.25. (B) $K_{\mathrm{A},n}$ histograms created from 10000 simulated single-molecule FRET time traces with the average time spent in the docked state $\tau_{\mathrm{high}} = 10$ s and in the undocked state $\tau_{\mathrm{zero}} = 10$ s. To calculate $K_{\mathrm{A},n}$ according to Equation 5.7, $c$ was set to 25 $10^{-9}$ mol $\mathrm{L}^{-1}$, which is equivalent to the IBS1* concentration described in the experimental section, and $N := i_{\max} = j_{\max} = 1, 3, 5, 10, 30$, or 90 exponentially distributed dwell times to determine $\langle t_{\mathrm{zero},n}\rangle$ and $\langle t_{\mathrm{high},n}\rangle$. The red curves correspond to a beta distribution fit according to Equation 5.27. (C) Cumulative probability distribution of the data shown in (B) were fitted to Equation 5.30. Fit results are shown in Figure 5.12.

lead to considerable narrowing of the simulated distributions, as would be expected for *FRET* time trajectories with a high number of transitions [397]. The ratio distribution of two gamma distributions has been described by Coelho and Mexia:

$$\begin{aligned} p(K_{\mathrm{A},n} &= \frac{[gamma]_{\tau_{\mathrm{high}}}}{[\mathrm{IBS1}^*[gamma]_{\tau_{\mathrm{zero}}}} = \frac{p_\Gamma(\langle t_{\mathrm{high},n,i_{\max}}\rangle)}{[\mathrm{IBS1}^*]p_\Gamma(\langle t_{\mathrm{zero},n,j_{\max}}\rangle)} \\ &= \frac{k^N}{B(N,N)}(1+RK_{\mathrm{A},n}^{-2N})K_{\mathrm{A},n}^{N-1} \end{aligned} \tag{5.26}$$

with $N := i_{\max} = j_{\max}$, and $R = 1/K_\mathrm{A}$. Note that $R \sim \tau_{\mathrm{zero}}/\tau_{\mathrm{high}} \sim k_{\mathrm{dock}}/k_{\mathrm{undock}}$. $B(N,N)$ is the so-called beta function [399]. To account for the variable number of events, an amplitude $A$ was introduced, expanding Equation 5.26 as follows:

$$p(K_{\mathrm{A},n} = A\frac{K_\mathrm{A}^{-N}}{B(N,N)}(1+\frac{K_{\mathrm{A},n}}{K_\mathrm{A}})^{-2N}K_{\mathrm{A},n}^{N-1} \tag{5.27}$$

Figure 5.11B shows the distributions of $n_{\max}$ = 10000 computed $K_{\mathrm{A},n}$ values with $i_{\max} = j_{\max}$ = 1, 3, 5, 10, 30 and 90 exponentially distributed dwell-times in the high and zero *FRET* state. Approximation of the simulated data using Equation 5.27 results in an excellent fit. Here, it is preferable to plot such distributions in a cumulative manner, because the result is not dependent on the bin size, a method routinely used in diffusion analysis [395, 400]. Figure 5.11C shows a cumulative probability distribution $P(K_{\mathrm{A},n})$ of the simulated data from Figure 5.11B normalised to 1. To the best of our knowledge it is, however, not possible to analytically describe this type of data, since there is no analytical solution to the integral over Equation 5.27.

### Approximation of intermolecular heterogeneity

Logistic functions are well known and widely applied in many fields, for example to describe a bimolecular association [401, 402]. Figure 5.11C demonstrates that cumulative probability distributions $P(K_{\mathrm{A},n})$ can be very well approximated using the logistic function as defined in Origin 9.0 (OriginLab Coorp., Northhampton, MA, USA):

$$p(K_{\mathrm{A},n} = \frac{A_1 - A_2}{1+(K_{\mathrm{A},n}/K_\mathrm{A})^p} + A_2 \tag{5.28}$$

Here, the parameter $p$ determines the steepness of the transition in $P(K_{\mathrm{A},n})$, *i.e.* the width of $p(K_{\mathrm{A},n})$, as described above. $A_1$ and $A_2$ are the lower and upper boundaries of the sigmoid. The relation between $p$ and $i_{\max}$ (and $j_{\max}$) is illustrated in Figure 5.12.

Upon normalisation ($A_1 = 0$ and $A_2 = 1$) and transformation, Equation 5.28 turns into:

$$\begin{aligned} P^{norm}(log(K_{\mathrm{A},n})) &= 1 - \frac{1}{1+(\frac{10^{log(K_{\mathrm{A},n})}}{10^{log(K_{\mathrm{A}})}})^{p}} = \frac{1}{1+(10^{log(K_{\mathrm{A},n})-log(K_{\mathrm{A}})})^{p}} \\ &= \frac{(10^{log(K_{\mathrm{A},n})-log(K_{\mathrm{A}})})^{p}}{1+(10^{log(K_{\mathrm{A},n})-log(K_{\mathrm{A}})})^{p}} \end{aligned} \tag{5.29}$$

Substitution of -log($K_{\mathrm{A}}$) by $pK_{\mathrm{A}}$ yields:

$$P^{\mathrm{norm}}(pK_{\mathrm{A},n}) = \frac{(10^{-pK_{\mathrm{A},n}+pK_{\mathrm{A}}})^{p}}{1+(10^{-pK_{\mathrm{A},n}+pK_{\mathrm{A}}})^{p}} \tag{5.30}$$

Please note that Equation 5.30 is equivalent to Equation 5.32, which is well-known to describe a simple acid-base equilibrium:

$$\mathrm{A}^{-} + \mathrm{H}^{+} \rightleftharpoons \mathrm{AH} \tag{5.31}$$

$$f = \frac{f_{\mathrm{A}} + f_{\mathrm{AH}} 10^{\mathrm{p}K_{\mathrm{AH}}-pH}}{1+10^{\mathrm{p}K_{\mathrm{AH}}-pH}} \tag{5.32}$$

where $f$ corresponds to the fraction of protonated HA (fully deprotonated: $f_{\mathrm{A}}$ = 0, fully protonated: $f_{\mathrm{AH}}$ = 1) in a normalised from, for example in chemical shift mapping using NMR [403]. Equations 5.30 and 5.32 are equivalent with exception of the parameter $p$. In a normal acid-base equilibrium, the slope of the transition is not variable ($p$ = 1), and does therefore not provide any information about the system. In turn, single-molecule FRET experiments have finite observation time, leading to cross-sample variability and p values greater than 1. Equation 5.30 describes a homogeneous system with one averaged $K_{\mathrm{A}}$ value. Description of a heterogeneous system is achieved by introducing an amplitude as and by describing $P^{\mathrm{norm}}(K_{\mathrm{A},n})$ as a

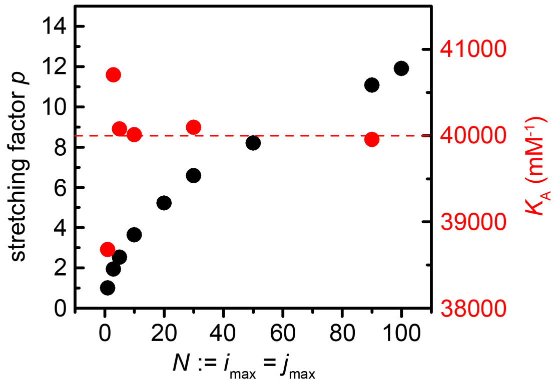


Figure 5.12: The relation between shape parameter $p$ and the number of dwell-times $N := i_{\max} = j_{\max}$ derived by fitting Equation 5.24 to the graphs shown in Figure 5.11. When $K_{\mathrm{A}}$ is set to 4 $10^{7}$ $\mathrm{M}^{-1}$, the simulations reveal that the smaller the number of dwell times during acquisition, the larger the deviation from $K_{\mathrm{A}}$.

linear combination of the total number $s_{\max}$ of components $K_{\mathrm{A},s}$:

$$P^{\mathrm{norm}}(K_{\mathrm{A},n}) = \sum_{s=1}^{s_{\max}} a_s \frac{1}{1+(\frac{K_{\mathrm{A},s}}{K_{\mathrm{A},n}})^p}; \sum_{s=1}^{s_{\max}} = 1 \tag{5.33}$$

$$P^{\mathrm{norm}}(log(K_{\mathrm{A},n})) = \sum_{s=1}^{s_{\max}} a_s \frac{(10^{-pK_{\mathrm{A},n}+pK_{\mathrm{A}}})^p}{1+(10^{log(K_{\mathrm{A},n})-log(K_{\mathrm{A},s})})^p}; \sum_{s=1}^{s_{\max}} = 1 \tag{5.34}$$

Approximation of the experimental data with $P^{\mathrm{norm}}(K_{\mathrm{A},n})$ according to Equation 5.33 requires the value of $p$ to be known. $p$ depends on the number of transitions between different *FRET* states ($i_{\max}$ and $j_{\max}$, Figure 5.12), which are usually not constant between individual time traces (Figure 5.1B). In our hands, $i_{\max}$ and $j_{\max}$ range between 2 and 7 for the majority of molecules with a median value of approximately 3.5. As depicted in Figure 5.11, the widths of $P(K_{\mathrm{A},n})$ distributions are rather similar in this range. Hence, approximation of the shape of $P(K_{\mathrm{A},n})$ with $p$ = 1.9 (corresponding to $i_{\max} = j_{\max}$ = 3.5) leads to good results.

### 5.3.4 Description of experimental $K_{\mathrm{A}}$ distributions

The formalism described above was used as a tool to extract information about the relative population of $K_{\mathrm{A},s}$ values within a certain range in the case of a heterogeneous system. For fitting the experimental distributions $P^{\mathrm{norm}}(\log(K_{\mathrm{A},n}))$ to Equation 5.34, a minimum of three components ($s_{\max}$ = 3) are required to describe the experimental data, demonstrating that finite observation times can account only for a small fraction of the $\log K_{\mathrm{A},n}$ distribution width (Figures 5.10C,D and 5.13) [224]. The fit divides the overall population of molecules into three subpopulations displaying distinctly different $\log K_{\mathrm{A},s}$ values (Table 5.5). As a control, heterogeneity of d3'EBS1*/IBS1* interaction in the presence of $M^{2+}$ ions was confirmed by titrating IBS1* to d3'EBS1* at 8 mM $Mg^{2+}$: Again, the fractions docked are more adequately represented when three $\log K_{\mathrm{A}}$ values are taken into account (Equation 5.34, $s_{\max}$ = 3, Table 5.6 and Figure 5.14). Since limited observation times do not allow for precise determination of $\log K_{\mathrm{A}}$ of type IIIa and IIIb molecules, these values were approximated with a $\log K_{\mathrm{A}}$ value of 10, respectively. Taken together, while d3'EBS1*/IBS1* docking behaves homogeneously in the absence of $M^{2+}$, the undocking kinetics are heterogeneous irrespective of the absence or presence of either $Mg^{2+}$ or $Ca^{2+}$. We accurately describe such a heterogeneous system by the here derived and generally applicable method, taking three subpopulations into account.

### 5.3.5 Metal ion binding sites in d3'EBS1* and d3'EBS1*/IBS1*

In order to rationalise the observed heterogeneity, we examined the recently solved solution structures of d3'EBS1* and d3'EBS1*/IBS1*, both containing several $Mg^{2+}$ binding sites as identified by chemical shift perturbation assays [50]. In the absence of IBS1*, the EBS1* loop is at least partly unstructured and flexible (Figure 5.15A). The $Mg^{2+}$ binding site located in the major groove of the d3' stem near the G4-C26 base pair remains unaffected by the absence or presence of IBS1*. A second binding site is present at the transition between the stem and the loop (A10, A20, and U9-G21 wobble pair). Upon IBS1* docking, EBS1* undergoes a substantial structural rearrangement to form a rigid structure (Figure 5.15B), thereby also altering $M^{2+}$ binding. The binding site at the loop-stem transition is slightly shifted from the G-U wobble towards the 5'-end of IBS1* and an additional binding site forms between EBS1* and IBS1* near the first two nucleotides of EBS1*, G13 and G14 (compare panels A and B in Figures refheterogeneity17 and 5.17). We determined the stability constants of $Mg^{2+}$ binding to each site in d3'EBS1* and d3'EBS1*/IBS1* from chemical shift perturbations [48], all being in the low millimolar range ($2.63 \leq \log K_A \leq 3.06$, Figure 5.15).

To pinpoint the nature of the $Mg^{2+}$-RNA interaction, we now carried out additional NMR titration studies with $Mn^{2+}$ and $[Co(NH_3)_6]^{3+}$. $Mn^{2+}$ has an ionic radius similar to $Mg^{2+}$ and can coordinate directly (inner-sphere) to nucleic acids, causing line broadening of resonances from protons nearby. $[Co(NH_3)_6]^{3+}$ is a kinetically stable octahedral complex used to mimic the

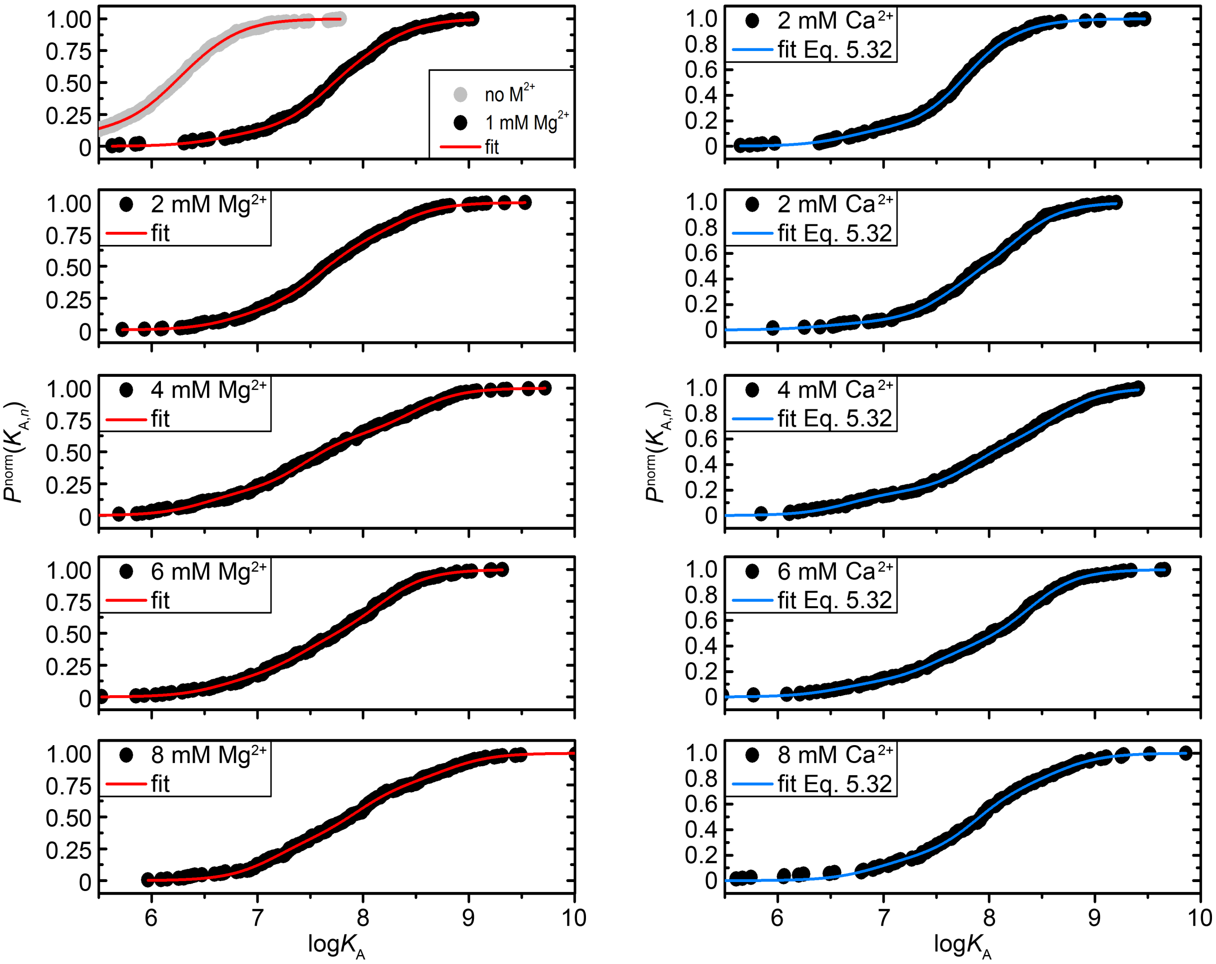


Figure 5.13: Normalised cumulative probability distributions of $K_{A,n}$ values determined from individual dynamic d3'EBS1* molecules (type I). Fits were performed using Equation 5.34 with $s_{\max}$ = 3 and $p$ = 1.9 (which corresponds to $i_{\max} = j_{\max}$ = 3.5). Fit results are shown in Table 5.5.

binding of $[\mathrm{Mg(H_2O)_6}]^{2+}$ (outer-sphere coordination) [405]. NOEs between the ammonia ligands and RNA protons directly reveal binding. The combination of $\mathrm{Mn}^{2+}$ and $[\mathrm{Co(NH_3)_6}]^{3+}$ binding thus reveals differences in outer and inner-sphere coordination of $\mathrm{M}^{2+}$ ions. We find that each site can be occupied by both species (Figures 5.16 and 5.17), but that the two loop binding sites show some preference for a specific type of coordination. $[\mathrm{Co(NH_3)_6}]^{3+}$ interacts very strongly with the closing G-U wobble in the d3'EBS1* stem but less pronounced in the presence of IBS1*, indicating that the $\mathrm{M}^{2+}$-ion docked here must be partially dehydrated for binding when EBS1*/IBS1* is formed (Figure 5.16). In the EBS1*/IBS1* duplex, both $[\mathrm{Co(NH_3)_6}]^{3+}$ and $\mathrm{Mn}^{2+}$ interact with G13 and G14. U62-U64 on IBS1* are markedly more affected by $[\mathrm{Co(NH_3)_6}]^{3+}$ than by $\mathrm{Mn}^{2+}$ (Figures 5.16B and 5.17B). One reason for this may be that the two species bind to EBS1*/IBS1* in a slightly different position as $[\mathrm{Co(NH_3)_6}]^{3+}$ is considerably bulkier than the $\mathrm{Mn}^{2+}$ ion. Importantly, the kinked structure of the loop (Figure 5.15B) may further restrict the accessibility of the phosphate moieties in the EBS1*/IBS1* major groove and of the 5'-end EBS1* nucleotides, which may be regarded as the strongest $\mathrm{M}^{2+}$ binding site [2]. In conclusion, two $\mathrm{M}^{2+}$ ions are directly involved in EBS1*/IBS1* binding, the $\mathrm{M}^{2+}$-RNA interactions being partly mediated by direct inner-sphere binding.

Table 5.5: Affinity constants of dynamic d3'EBS1*-IBS1* molecules: $K_{\mathrm{A},1}$, $K_{\mathrm{A},2}$, $K_{\mathrm{A},3}$, and their corresponding amplitudes $a_1$, $a_2$, and $a_3$ in the presence of $\mathrm{M}^{2+}$ obtained by fitting Equation 5.34 to the experimental distributions $P(K_{\mathrm{A},n})$ shown in Figure 5.13. The last column lists the averaged $\log K_{\mathrm{A,av}}$: $\log \sum_{s=1}^{s_{\max}} a_s\, 10^{logK_{\mathrm{A},s}}$.

| $c(\mathrm{M}^{2+}$ | $\log K_{\mathrm{A},1}$ | $\log K_{\mathrm{A},2}$ | $\log K_{\mathrm{A},3}$ | $p$ | $a_1$ | $a_2$ | $a_3$ | $\log K_{\mathrm{A,av}}$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 5.20 | 6.17 | 6.60 | 1.9 | 0.14 | 0.59 | 0.27 | 6.29 |
| 1 mM $\mathrm{Mg}^{2+}$ | 6.64 | 7.62 | 8.18 | 1.9 | 0.11 | 0.58 | 0.31 | 6.29 |
| 1 mM $\mathrm{Ca}^{2+}$ | 6.67 | 7.77 | 8.33 | 1.9 | 0.14 | 0.80 | 0.06 | 7.78 |
| 2 mM $\mathrm{Mg}^{2+}$ | 6.80 | 7.58 | 8.33 | 1.9 | 0.16 | 0.55 | 0.29 | 7.92 |
| 2 mM $\mathrm{Ca}^{2+}$ | 6.43 | 7.60 | 8.27 | 1.9 | 0.06 | 0.41 | 0.54 | 8.07 |
| 4 mM $\mathrm{Mg}^{2+}$ | 6.51 | 7.46 | 8.48 | 1.9 | 0.19 | 0.45 | 0.36 | 8.09 |
| 4 mM $\mathrm{Ca}^{2+}$ | 6.67 | 7.83 | 8.64 | 1.9 | 0.19 | 0.41 | 0.41 | 8.31 |
| 6 mM $\mathrm{Mg}^{2+}$ | 6.66 | 7.41 | 8.17 | 1.9 | 0.16 | 0.34 | 0.50 | 7.92 |
| 6 mM $\mathrm{Ca}^{2+}$ | 6.57 | 7.49 | 8.39 | 1.9 | 0.12 | 0.29 | 0.59 | 8.19 |
| 8 mM $\mathrm{Mg}^{2+}$ | 7.15 | 7.96 | 8.80 | 1.9 | 0.33 | 0.44 | 0.23 | 8.28 |
| 8 mM $\mathrm{Ca}^{2+}$ | 6.93 | 7.87 | 8.66 | 1.9 | 0.18 | 0.57 | 0.24 | 8.19 |

Table 5.6: Affinity constants of dynamic d3'EBS1*-IBS1* molecules: $K_{\mathrm{A},1}$, $K_{\mathrm{A},2}$, $K_{\mathrm{A},3}$, and their corresponding amplitudes $a_1$, $a_2$, and $a_3$ in the presence of a constant concentration of 8 mM $\mathrm{Mg}^{2+}$ or $\mathrm{Ca}^{2+}$ and increasing IBS1* concentration. The last column lists the averaged $\log K_{\mathrm{A,av}}$: $\log \sum_{s=1}^{s_{\max}} a_s\, 10^{logK_{\mathrm{A},s}}$.

| | $c(\mathrm{M}^{2+})$ | $\log K_{\mathrm{A},1}$ | $\log K_{\mathrm{A},2}$ | $\log K_{\mathrm{A},3}$ | $p$ | $a_1$ | $a_2$ | $a_3$ | $\log K_{\mathrm{A,av}}$ |
|---|---|---|---|---|---|---|---|---|---|
| 8 mM $\mathrm{Mg}^{2+}$ | 1 nM | 5.03 | 7.12 | 8.01 | 1.9 | 0.12 | 0.29 | 0.59 | 7.81 |
| | 5 nM | 6.46 | 7.75 | 8.82 | 1.9 | 0.29 | 0.27 | 0.44 | 8.48 |
| | 10 nM | 4.78 | 7.29 | 7.96 | 1.9 | 0.07 | 0.35 | 0.58 | 7.78 |
| | 25 nM | 4.62 | 7.34 | 8.28 | 1.9 | 0.18 | 0.13 | 0.69 | 8.13 |
| | 40 nM | 5.28 | 7.85 | 8.77 | 1.9 | 0.13 | 0.42 | 0.45 | 8.47 |
| 8 mM $\mathrm{Ca}^{2+}$ | 1 nM | 4.44 | 7.15 | 7.78 | 1.9 | 0.17 | 0.32 | 0.51 | 7.54 |
| | 5 nM | 4.47 | 7.45 | 8.21 | 1.9 | 0.09 | 0.41 | 0.50 | 7.78 |
| | 10 nM | 4.35 | 7.56 | 8.28 | 1.9 | 0.11 | 0.30 | 0.59 | 8.09 |
| | 25 nM | 4.54 | 8.18 | 8.84 | 1.9 | 0.11 | 0.49 | 0.40 | 8.55 |
| | 40 nM | 5.22 | 7.97 | 8.89 | 1.9 | 0.16 | 0.43 | 0.41 | 8.55 |

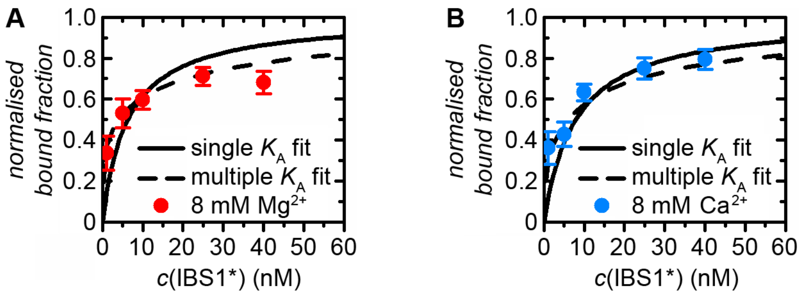


Figure 5.14: IBS1*-dependent average docked fraction as determined from $>200$ smFRET time traces recorded at (A) 8 mM $\mathrm{Mg}^{2+}$ and (B) 8 mM $\mathrm{Ca}^{2+}$ (type IIIa molecules were omitted). Solid lines correspond to fit of the experimental data to Equation 5.21 ($l_{\mathrm{total}} = 1$), assuming a single average $\log K_{\mathrm{A}}$ value. Dashed lines correspond to fits using Equation 5.21 ($\mathrm{l}_{total} = 4$), assuming the occurrence of 4 subpopulations displaying distinct $\log K_{\mathrm{A}}$ values. Amplitudes ($a_1$, $a_2$, and $a_3$) and the corresponding $\log K_{\mathrm{A}}$ values describing dynamic time traces were taken from Table 5.6, while static docked molecules (type IIIb) were approximated with $\log K_{\mathrm{A},4} = 10$. The sum of fractions $f_1,...,f_4$ was normalised to 1. Error bars correspond to $2\sigma$bootstrap [166]. The sum of the squared residuals for single versus multiple $\log K_A$ fits are 0.089 vs. 0.010 for $\mathrm{Mg}^{2+}$, and 0.071 vs. 0.013 for $\mathrm{Ca}^{2+}$, respectively. This clearly demonstrates that multiple $\log K_{\mathrm{A}}$ fits describe the experimental data better than the single $\log K_{\mathrm{A}}$ fits.

## 5.4 Discussion

This combined single-molecule NMR study addresses the influence of the physiologically relevant metal ions $Mg^{2+}$ and $Ca^{2+}$ on a basic RNA-RNA interaction, the d3'EBS1*/IBS1* splice-site recognition from group II intron ribozymes. In agreement with previous studies, already small amounts of $M^{2+}$ ions strongly shift the thermodynamic equilibrium towards the docked state [388]. Strand association in the absence of $M^{2+}$ behaves homogeneous: The presence of $M^{2+}$ renders also this step heterogeneous as is dissociation under all conditions. NMR experiments show that two metal ions are located in proximity to the d3'EBS1*/IBS1* duplex. Hence, the occurrence of non-conventional RNA-RNA interaction kinetics coincides with the presence of specific $M^{2+}$ binding sites at the interaction site of the two strands.

Heterogeneous undocking behavior in the absence of $M^{2+}$ ions suggests a rugged energy landscape with more than one thermodynamically stable d3'EBS1*/IBS1* conformation [226]. It is likely that $M^{2+}$ ions further stabilise (local) energy minima, leading to substructures with similar global folds, but differences on the microscopic level, as well as a broad distribution of docking/undocking kinetics. As the metal ion binding pockets are occupied only approximately half of the time at the range of $M^{2+}$ concentrations assessed in our smFRET experiments, we propose that conformational heterogeneity is for the most part due to differences in binding site occupation (Figure 5.18). As substructures are expected to differ in topology, they are highly likely to display different interstrand affinities and/or association/dissociation kinetics [2]. Importantly, the exchange rate of the $M^{2+}$ ions, $\sim 10^6$ and $\sim 10^8$ $s^{-1}$ for hexahydrated $Mg^{2+}$ and $Ca^{2+}$, will be decreased by several orders of magnitude through multiple RNA-$M^{2+}$ interac-

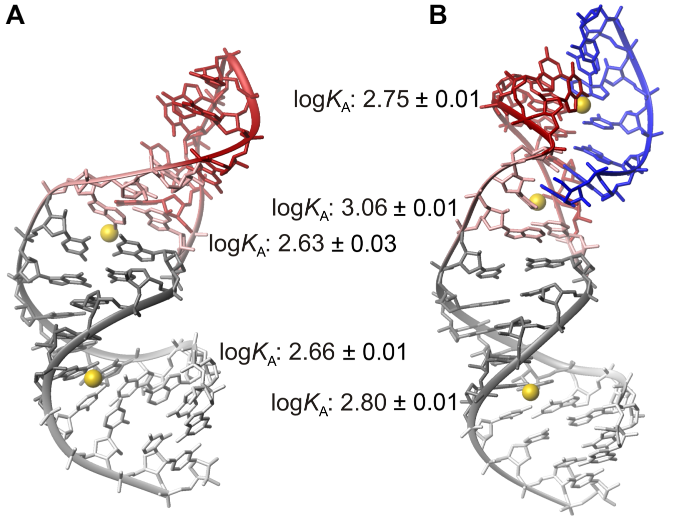


Figure 5.15: Solution structure and proposed metal ion binding sites of (A) d3'EBS1* and (B) d3'EBS1*/IBS1*. The lowest energy structures from an ensemble of 20 is shown. Yellow spheres mark possible binding positions of $Mg^{2+}$ for which an estimate of the affinity constant of $Mg^{2+}$ binding to the RNA is given [50]. The figure was prepared using MOLMOL from pdb entries 2M24 and 2M23 and the colouring scheme is consistent with Figure 5.1A [404].

tions and inner-sphere coordination as observed here by NMR [66, 105]. Metal ion exchange rates in RNA binding pockets with comparable affinities have been estimated to take place at $10^3$ $s^{-1}$ [406, 407]. As this timescale reaches the time resolution attainable with CCD cameras typically used in smFRET experiments, heterogeneous occupation of the metal ion binding sites is expected not to be completely averaged out [161]. As a consequence, the heterogeneity of d3'EBS1*/IBS1* affinity most likely results from the interplay between the structural rearrangements undergone upon helix formation and the dynamics of relevant $Mg^{2+}$-binding.

## 5.5 Conclusion

Conformational changes in RNA have often been observed to display multiexponential kinetics, but discord persists with regard to the origin of the phenomenon [195, 197, 198, 199, 200, 213, 219, 224]. However, when care is taken to rule out experimental artifacts, it is generally agreed upon that heterogeneity observed in single molecule experiments is a real phenomenon

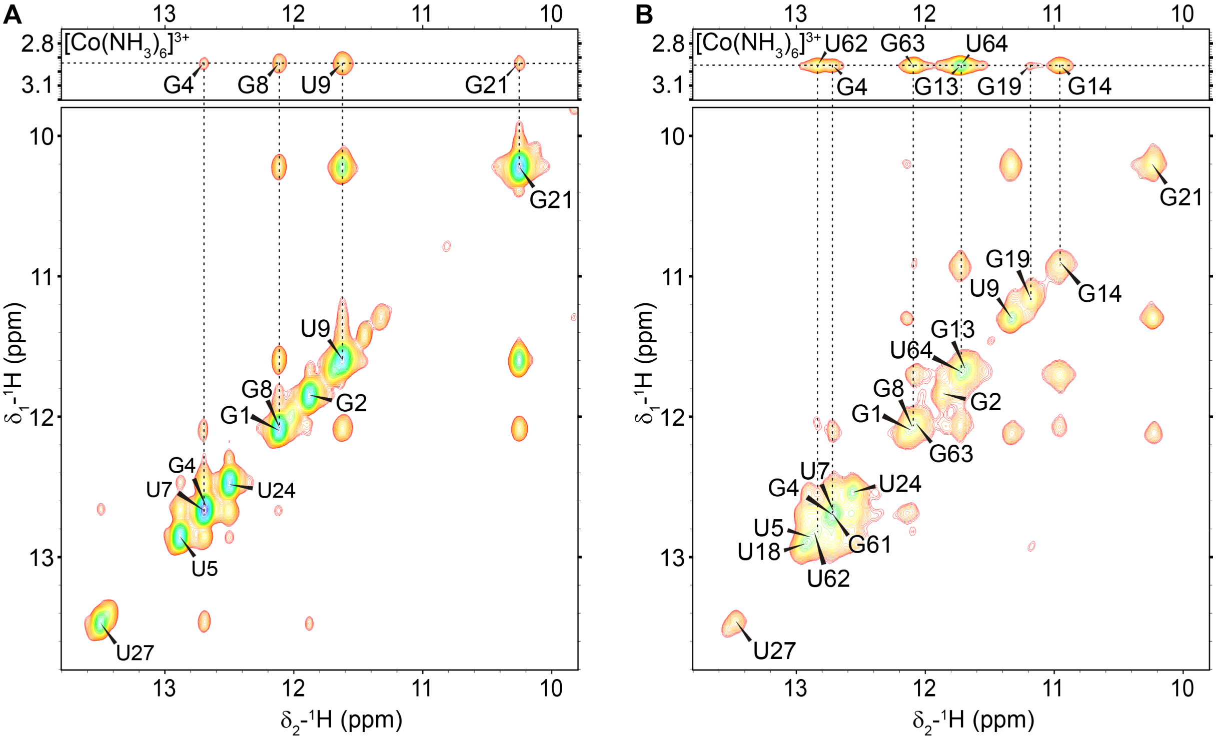


Figure 5.16: Cobalt hexamine binding to (A) d3'EBS1* and (B) d3'EBS1*/IBS1* as observed by NOE cross peaks in a [$^1$H,$^1$H]-NOESY of the imino protons in uracil and guanine. Cobalt hexamine binds in close proximity to the U9-G21 wobble pair in d3'EBS1*. In d3'EBS1*, this base pair seems not to bind cobalt hexamine, but strong correlations to bases at/near the 5'-end of EBS1* (G13, G14 and U64) are observed. Spectra were recorded in 90 % $H_2O$/10 % $D_2O$ at 278 K in 10 μM EDTA and 10 mM KCl, 2.5 mM $[Co(NH_3)_6]^{3+}$ (d3'EBS1*) and 110 mM KCl, 1.5 mM $[Co(NH_3)_6]^{3+}$ (d3'EBS1*/IBS1*). The RNA concentration was consistently set to 0.55 mM.

[225, 226]. A direct link between intermolecular heterogeneity and the availability of metal ions has been observed for RNAse P RNA and the human telomeric G-quadruplex [230, 408]. Importantly, it has been speculated (but never been proven) that incomplete occupation of $M^{n+}$ binding sites induces microscopic heterogeneity between individual molecules, leading to a pronounced memory effect [408]. Here, by combining single-molecule techniques and NMR to investigate a basic RNA-RNA interaction within a group II intron, we provide direct evi-

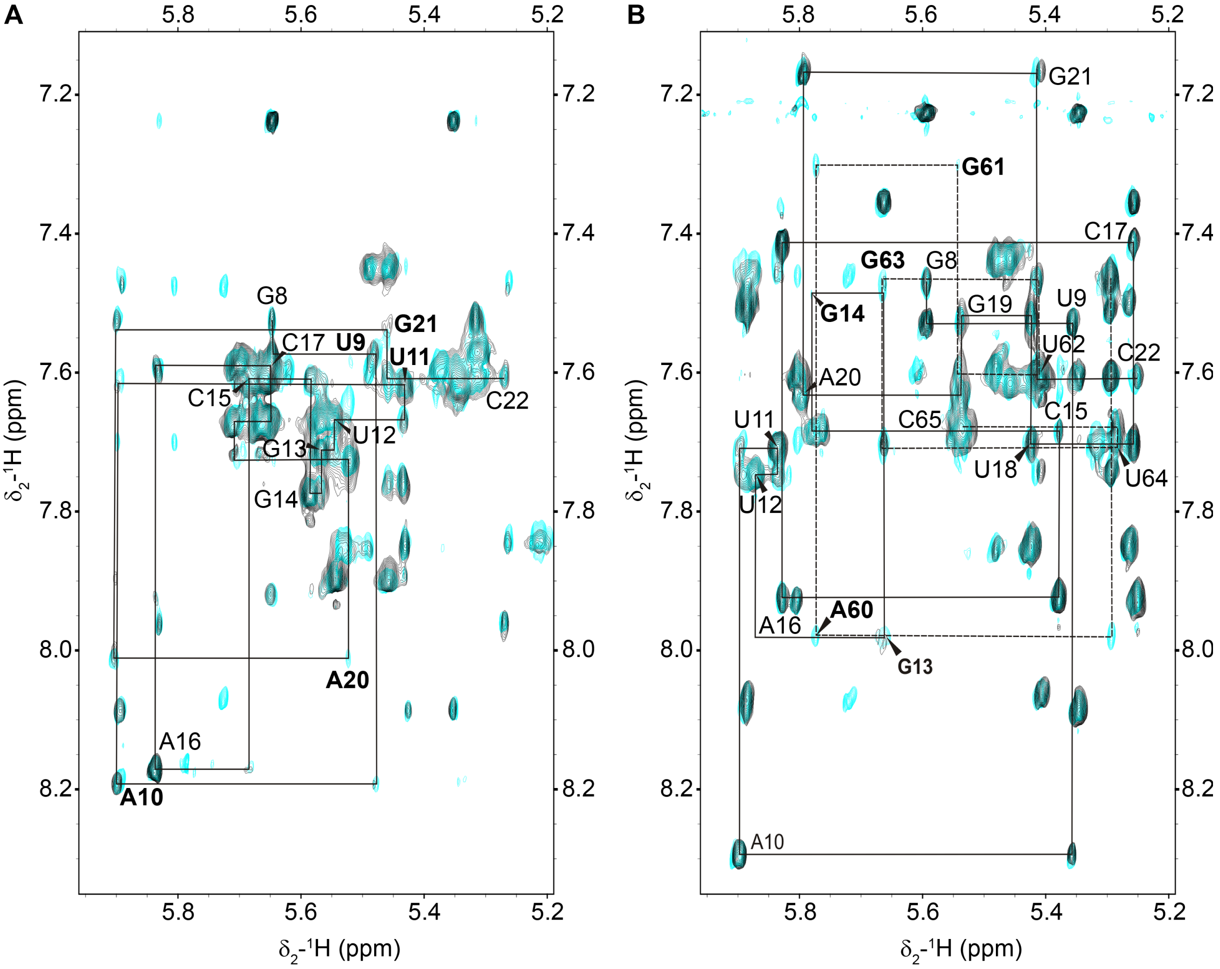


Figure 5.17: $Mn^{2+}$ binding to (A) d3'EBS1* and (B) d3'EBS1*-IBS1* as observed by paramagnetic line broadening in [$^1$H,$^1$H]-NOESY spectra of the non-exchangeable protons. Shown is a superposition of the spectra recorded in the absence of $Mn^{2+}$ (cyan) or presence of 150 μM $Mn^{2+}$ or 270 μM $Mn^{2+}$ (black ; d3'EBS1* or d3'EBS1*-IBS1*, respectively). Cross peaks that have been broadened to base line at the given concentration of $Mn^{2+}$ appear in cyan only. The sequential connectivities between H1' and H6/H8 protons of residues G8 and C22 are indicated as solid lines. Dashed lines visualise connectivities between residues C59-C65 of IBS1*. Residues mainly affected by line broadening are highlighted using bold face type. In the loop of d3'EBS1*, cross peaks between U9, A10 and U11 as well as cross peaks of A20 and G21 are most affected by broadening. In turn, cross peaks of G13, G14, A60 and G63 are the most affected peaks of the loop region within d3'EBS1*/IBS1*. Spectra were recorded in 100 % $D_2O$ (d3'EBS1*: 0.54 mM RNA, 10 mM KCl, 10 μM EDTA, 293 K; d3'EBS1*-IBS1*: 0.6 mM RNA, 110 mM KCl, 10 μM EDTA, 298 K).

dence for this hypothesis. As M$^{n+}$ binding is inextricably linked with nucleic acid structure formation and numerous metal ion binding sites are not fully occupied at physiological cation concentrations, one may speculate that heterogeneity is a general property of functional nucleic acids [66, 406, 407]. The importance of heterogeneity is increasingly appreciated and led to the recent revision of fundamental laws of enzymology (Michaelis-Menten equation) and allostery (Hill equation) [225, 409]. Our here presented method provides a straightforward way to approximate, characterise, and compare such heterogeneities. Similarly, as molecules are generally regarded as homogeneous in bulk studies, future work is anticipated to extend the methodologies typically used to analyse ensemble data. Furthermore, it will be of fundamental interest to find out whether unconventional kinetics are present in vivo and whether cofactors exist that control and/or modulate this effect.

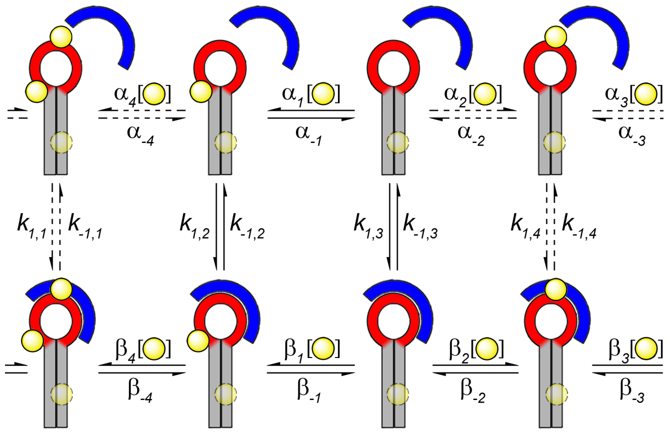


Figure 5.18: Working model of cation-dependent d3'EBS1*/IBS1* interaction kinetics. Several d3'EBS1* conformers differing in metal ion binding site occupation co-exist at physiological Mg$^{2+}$ or Ca$^{2+}$ concentrations and they interconvert at the cation exchange rates ($\alpha$, $\beta$). Conformers further differ with regard to IBS1* docking and undocking kinetics ($k_1$, $k_{-1}$). The distant cation binding site located near the G4-C26 base pair is assumed to have little influence on docking/undocking rates. Dashed lines indicate conversions that were not detected experimentally.

## Chapter 6

# Kinetic subpopulations detected by single-molecule spectroscopy - fundamental property of functional nucleic acids or experimental artefact?

Sebastian L.B. König, Danny Kowerko, Roland K.O. Sigel

Institute of Inorganic Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

**ABSTRACT:** Single-molecule spectroscopy allows the direct observation of conformational dynamics in individual biomolecules. Here, we describe how single-molecule Förster resonance energy transfer (smFRET) reveales heterogeneous kinetics in the EBS1*/IBS1* interaction, two RNA sequences that play an important role in group II intron mediated self-cleavage. Further examples of dynamic heterogeneity in functional nucleic acids are provided and the possible origins of this phenomenon are discussed.

**Keywords:** group II introns, splicing, smFRET, heterogeneity, memory effect

Published in *CHIMIA* 2013, **67**, 240-243, invited article.

## 6.1 Introduction

Fluorescence-based methods have experienced unprecedented popularity to address biological questions. Detection of Förster resonance energy transfer (FRET) between a single pair of fluorophores attached to a biopolymer is of particular interest, because it provides a handle on measuring stochastic, time-dependent fluctuations in its structure [169]. FRET is the distance-dependent, non-radiative energy transfer between two dipoles, typically the transition dipole moments of two fluorophores that are referred to as donor and acceptor [28]. By estimating the interdye distance in real time, single-molecule FRET (smFRET) therefore allows for the detection of intermediates along the folding pathway of biologically relevant molecules, as well as misfolded structures. As a consequence, it has been frequently used to study folding and function of catalytically active RNA molecules (ribozymes), regulatory elements of translation (riboswitches), and functional DNAs [170, 410].

In a typical smFRET experiment, nucleic acid samples are labeled with a fluorophore pair and a biotin moiety for subsequent immobilisation, even though smFRET studies can also be performed in solution, an approach that is routinely used to characterise proteins [290, 411]. Suitable dyes for FRET must be (i) bright (high quantum yield, high extinction coefficient $\varepsilon$), (ii) photostable (stable emission intensity, no blinking), (iii) small (minimal perturbation of the system under study), (iv) water-soluble and (v) easily attachable to the biomolecule of interest [141]. Moreover, the fluorophore pair must display spectral overlap for FRET to occur. Cyanine dyes are the dominant choice as FRET probes in nucleic acid biophysics, but other dyes have also been successfully implemented [412]. smFRET measurements are then carried out using either a confocal microscope or a total internal reflection fluorescence (TIRF) system (Figure 6.1). Raw data are corrected for background noise, leakage of donor emission in to the acceptor channel and direct excitation of the acceptor, so-called crosstalk. Finally, the corrected emission intensities are used to calculate *FRET* efficiency in order to probe intramolecular motion and intermolecular interaction. For further information on the implementation of smFRET and data analysis we direct the reader to the following excellent reviews [141, 390].

The yeast Group II intron *Sc*.ai5$\gamma$ is one example where smFRET has been successfully employed to study ribozyme dynamics [87, 204, 205]. These catalytic RNAs are several hundreds of base pairs in length and catalyse their own excision from the nascent messenger RNA molecule [204]. Recognition of the correct cleavage sites, *i.e.* the splice sites, is brought about by interaction of exon-binding sequences and intron-binding sequence (EBS1-3, IBS1-3) [388]. Here, a simple model system to study the influence of divalent cations on 5’ splice site formation is proposed (Figure 6.2A), followed by reporting and discussing heterogeneity observed with regard to its docking/undocking behavior [413].

## 6.2 Materials and methods

### 6.2.1 Single-molecule FRET imaging

The experimental design (Figure 6.2A) involved surface immobilisation of the Cy3-labeled d3'EBS1* hairpin within a home-built microfluidic chamber via a streptavidin-biotin linkage [170, 388]. smFRET movies were recorded using a TIRF microscope under the following imaging conditions: 100 mM KCl, 50 mM MOPS, 1 % D-glucose (w/V), 1x enzymatic oxygen scavenger solution (to prolongate dye lifetime), and 1 mM Trolox (for stable emission intensities), a protocol modified from Selvin and Ha [412]. Furthermore, the imaging buffer contained 25 nM Cy5-labelled IBS1* and variable $MgCl_2$/$CaCl_2$ concentration between 0 and 8 mM. Single-molecule clips were recorded over 6 minutes at 10 frames/s using an Andor 897 CCD camera (Lot Oriel, Romanel-sur-Morges, Switzerland). Oligonucleotides were ordered from Microsynth (Balgach, Switzerland) and chemicals were purchased *purissum* grade from Sigma-Aldrichs (Buchs, Switzerland).

### 6.2.2 Data analysis

Single-molecule clips analysed using a home-written Matlab-based software called SIRA (SIf Reader and Analyser), an approach detailled elsewhere [390]. Briefly, single fluorophore time

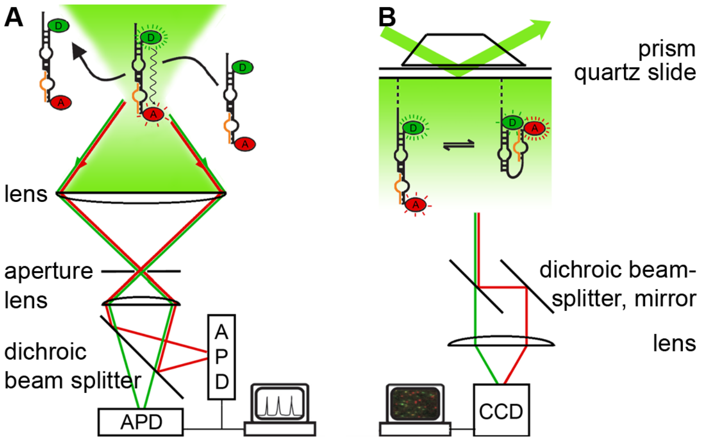


Figure 6.1: Outline of single-molecule FRET microscopy systems. (A) Confocal microscopy. The donor dye (*D*) is excited when the biomolecule traverses the focal volume (green area). FRET occurs when the acceptor dye (*A*) is sufficiently close in space. Out-of-focus signal is rejected by an aperture, while in-focus bursts of fluorescence are detected upon spectral separation using avalanche photodiodes (APDs). It should be noted that photon multiplier tubes (PMTs) have also been used for detection, as well as the recently developed hybrid PMTs that decrease detection artefacts (*afterpulsing*, Picoquant GmbH, Berlin, Germany, personal communication). (B) TIRF microscopy. The incident laser beam reaches the surface at an angle that is larger than the critical angle, which leads to its total reflection and creates an evanescent wave reducing the excited volume to a thin sheet (green). Dye emission of surface-tethered biomolecules is spectrally separated and projected side-by-side onto a CCD camera, which allows to simultaneously observe hundreds of molecules over several minutes. Figure modified from [169].

trajectories were corrected for background noise and crosstalk (*vide supra*) prior to manual selection for anticorrelated emission profiles. Subsequently, the apparent *FRET* efficiency was calculated from corrected fluorophore emission time traces according to equation 6.1,

$$FRET = \frac{PC^{\text{Cy3 exc}}_{\text{Cy5 em}}}{PC^{\text{Cy3 exc}}_{\text{Cy3 em}} + PC^{\text{Cy3 exc}}_{\text{Cy5 em}}} \tag{6.1}$$

where *PC* refers to the photon counts due to Cy3 or Cy5 emission upon Cy3 excitation, respectively.

## 6.3 Results

Single-fluorophore emission intensities fluctuated in a FRET-typical anticorrelated fashion (Figure 6.2B,C), followed by a sudden drop in emission to zero (not shown). Such single-step photobleaching is important to rule out the prevalence of doubly-labeled RNA molecules. *FRET* time traces revealed the occurrence of two *FRET* states that were assigned to the docked (high *FRET*) and the undocked state (no *FRET*), as FRET can only occur when Cy3 and Cy5 are sufficiently close in space, typically in the range of 30 - 80  depending on the FRET pair.

Interestingly, pronounced heterogeneity with regard to docking/undocking kinetics was observed in the presence of divalent cations (Figure 6.2). Conventional approaches to determine rate constants (dwell time analysis) required three docking and three undocking rate constants to describe the experimental data recorded in the presence of divalent metal ions [219]. Fur-

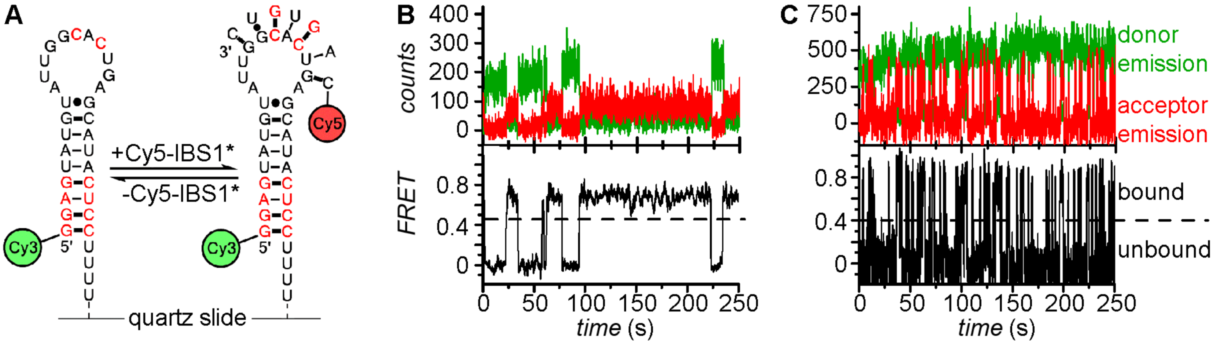


Figure 6.2: Experimental design and representative data. (A) The d3'EBS1* hairpin is Cy3-labeled and immobilised on the surface of a quartz slide. Docking of a Cy5-labeled IBS1* oligonucleotide is accompanied by a burst of Cy5 fluorescence and a decrease in Cy3 emission due to FRET. A: adenine, U: uracil, G: guanine, C: cytosine, interconnected via a phosphosugar backbone. Non-native bases are shown in red [388]. (B,C) Two representative time trajectories recorded in the presence of 8 mM $CaCl_2$ show clear differences in docking/undocking kinetics between individual molecules. Upper graphs: fluorophore emission over time reveals anticorrelated changes typical for FRET. Lower graphs: *FRET* over time shows fluctuations between a zero *FRET* state (undocked) and a high *FRET* state (docked). Figure modified from [413].

thermore, rate constants associated with docking and undocking could not be unambiguously assigned to each other, which precluded the existence of three discrete kinetic subpopulations. Rather, rate constants seem to be continuously distributed over several orders of magnitude (Figure 6.3), which suggests the coexistence of numerous (sub)structures.

Single-molecule studies rely on the principle of ergodicity, according to which the properties of billions of molecules can be reliably predicted by observing only a small number of them (here: roughly 200) [355]. The averaged $\Delta G^{\circ}$ values associated with IBS1* docking determined from single-molecule time trajectories were found to be in good agreement with the results of UV/vis thermal melting studies *et al.* [17]. These findings suggest that the complex docking/undocking kinetics observed in the smFRET experiments are not an experimental artefact. Consequently, a novel method to describe the distribution of equilibrium constants was proposed and applied. As a comprehensive description would go beyond the scope of this article, the authors direct the interested reader to the original article (see previous Chapter and ref. [413]).

## 6.4 Discussion

At first, the complex kinetic behavior observed for d3'EBS1*/IBS1* interaction may seem somewhat surprising given the simplicity of the experimental system. However, numerous examples of kinetic heterogeneities in nucleic acid folding have been precedented in single-molecule studies. Heterogeneity may be classified as static or dynamic, where static heterogeneity refers to the coexistence of stable subpopulations that do not interconvert (*memory effect*), while dynamically heterogeneous subpopulations interconvert during the time of observation [414].

Heterogeneous folding kinetics have first been described in the hairpin ribozyme, which catalyses cleavage and ligation reactions in the replication cycle of the tobacco ringspot virus [415]. In an analogous fashion to d3'ESB1*/IBS1*, the hairpin ribozyme alternates between a docked

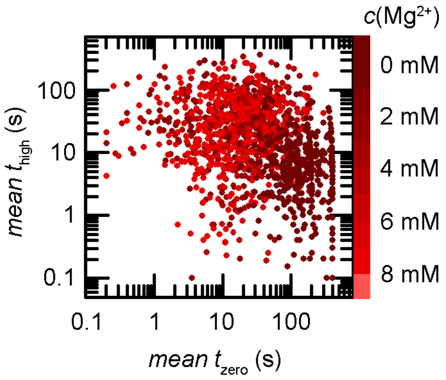


Figure 6.3: Kinetic heterogeneity of d3'EBS1*/IBS1* docking. The mean times spent in the docked ($t_{\text{high}}$) and in the undocked state ($t_{\text{zero}}$) are shown for individual EBS1* molecules at different $\text{Mg}^{2+}$ concentrations. Figure modified from [413].

and an undocked conformation and dynamics directly correlate with catalytic activity. While heterogeneity in cleavage kinetics had already been observed in bulk studies using a *minimal ribozyme*, *i.e.* a shorter version that is still catalytically active, Zhuang *et al.* could directly observe heterogeneity in structural dynamics between single hairpin ribozymes performing smFRET studies [195, 415]. While docking could be described with a single rate constant, undocking required a total number of four exponential fit functions to satisfactorily describe the experimental data and the resulting rate constants spanned four orders of magnitude [195]. In addition, individual ribozymes were observed to dissociate with similar rate constants over long periods with less than 5 % of interconversion within 3 hours (static heterogeneity) [195]. In a follow-up study, Ditzler and co-workers observed that the hairpin ribozyme migrates as two discrete bands on a non-denaturing gel electrophoretic shift assay, which differ in kinetics in subsequent smFRET experiments, thus confirming the longevity of the effect [198]. Please refer to Section 3.3.3 for an account of smFRET studies performed with the hairpin ribozyme, their findings, as well as controversy associated with its kinetic heterogeneity. Numerous examples of heterogeneous kinetics have followed and they are summarised in Table 6.1. Please refer to [163] for an excellent review on this topic.

Table 6.1: Further examples of kinetic heterogeneity of biologically relevant nucleic acids as revealed by single-molecule spectroscopy. This list is not intended to be exhaustive.

| Biological model | Key findings | References |
|---|---|---|
| 5' splice site in group II introns | 3 docking rate constants, 3 undocking rate constants in the presence of $Mg^{2+}$ or $Ca^{2+}$ | [390] |
| Minimal hairpin ribozyme | 1 docking rate constant, 4 undocking rate constants, static undocking heterogeneity | [195, 198] |
| Native hairpin ribozyme | Undocking rate constants span 3 orders of magnitude at low [$Mg^{2+}$], static undocking heterogeneity | [199] |
| *Tetrahymena* group I intron | Individual equilibrium constants vary by a factor of up to 300, $\geq$6 kinetic subpopulations, interconversion upon removal and re-addition of $Mg^{2+}$ | [219, 224] |
| Holliday junction | Partitioning of experimental data into five kinetic subspecies, interconversion upon removal and re-addition of $Mg^{2+}$ | [387] |
| *Human* telomeric G-quadruplex | Kinetic subspecies display either slow (min) or fast (s) structural rearrangements, interconversion between kinetic subgroups possible | [191] |

### 6.4.1 About the origins of heterogeneous kinetics

The origins of heterogeneous kinetics are likely to be diverse and they are subject of ongoing debate. One phenomenon known to bias smFRET studies is fluorophore photophysics, especially intersystem crossing to long-lived dark states leading to so-called blinking, as well as fluctuations in emission intensity [146]. For some cyanine dyes, the occurrence of undesired photophysical events can be greatly reduced by enzymatically removing molecular oxygen and upon addition of the vitamin E analogue Trolox [150]. Nevertheless, blinking, intensity fluctuations and spectral changes appear to be a universal property of fluorophores and can be more or less pronounced depending on the dyes chosen and their molecular environment [146]. Therefore, the experimental design must be chosen carefully in order to clearly differentiate between changes in *FRET* caused by molecular motion and those caused by dye photophysics, which could otherwise contribute to the kinetic analysis. Along these lines, appropriate control experiments are also invaluable (Figures 5.5 and 12.10, as well as Figures 12.4, 12.5, 12.6, 12.7, 12.8 and 12.9 in the Appendix).

Probably the most fundamental feature shared by all functional nucleic acids is the fact that a precise fold must be adopted for functionality to be achieved [23]. As there is an almost infinite number of potentially available conformations (especially for long sequences), biomacromolecules may also adopt misfolded structures corresponding to local minima in the free energy landscape [23]. Such kinetic traps are resolved in the living cell, for example by proteins, but they may accumulate in the artificial environment of the single-molecule studies that are frequently carried out under non-physiological conditions [28]. It is believed that kinetic heterogeneity is caused by the co-existence of several structures and a number of culprits have been proposed to this end (list adapted from [163]):

#### Surface tethering

Immobilisation of biomolecules might alter their structure and hence their functionality. In fact, surface-attachment has been shown to inactivate proteins [416]. Nevertheless, this does not seem to hold true for RNA, as indirectly immobilised hairpin ribozymes encapsulated within lipid vesicles have been shown to behave like directly surface-tethered molecules [200].

#### Chemical differences

Ditzler and co-workers excluded differences in molecular structure arising from chemical synthesis using high-resolution mass spectrometry, though this does not rule out mass-neutral modifications, for example structural isomers, diastereomers, and enantiomers [198]. Exposure to UV-light during the purification process (*shadowing*), as well as heating have recently been demonstrated to irreversibly alter the chemically structure of RNA and thus leading to a broadened free energy landscape [226]. Hence, care must be taken during RNA purification and annealing steps.

### Metal ion binding

Due to the polyanionic nature of the phosphate-sugar backbone, formation of nucleic acid structures is inextricably linked to the binding of metal ions [2]. Here, magnesium ions are of particular interest because they are known to be specifically involved in folding and catalysis and owing to their bio-availability [2]. However, compared to the affinity of metal ions to their specific binding sites in proteins ($K_\mathrm{D} \leq 10^{-6}$ M for surface-bound cations, $K_D \geq 10^{-7}$ M for *chaperoned* metal ions), their affinity to nucleic acid binding partners is lower by several orders of magnitude ($10^{-4}$ M $\leq K_\mathrm{D} \leq 10^{-2}$ M) [66, 417]. Metal ion binding to nucleic acids is consequently of dynamic nature and the simultaneous occurrence of substructures with different cation binding patterns would be expected. Along this line, Solomatin *et al.* demonstrated that interconversion between kinetic subspecies observed for the *Tetrahymena* group I intron could be greatly accelerated in $Mg^{2+}$ pulse experiments ($Mg^{2+}$ addition - removal - re-addition) [224]. Similar observations have recently been reported for the Holliday junction [387]. In turn, cation-dependent experiments performed with the hairpin ribozyme suggest that the role of divalent cations is to stabilise the transition state by forming contacts that do not exist in the native conformation ($\phi$ value analysis) [196, 197, 221]. Heterogeneities observed in group I introns and Holliday junction folding may therefore originate from differences in metal ion binding in the catalytically active fold, while this does not seem to hold true for the hairpin ribozyme. Through a combination of smFRET and NMR, direct evidence for heterogeneous metal ion binding and kinetic heterogeneity was provided recently [413].

### Absence of molecular crowding

Macromolecules, especially proteins, occur at high concentrations inside a living cell. In fact, their concentration is so high that typically 5 to 40 % of the total cell volume is physically occupied by these molecules [380]. As a consequence, such crowding conditions will not only favor any reaction leading to increased volume (formation of compact conformations, aggregates, etc.), they will also reduce diffusion coefficients by a factor of up to 10 through increased viscosity (Stokes-Einstein equation) [380]. Also, crowding agent will alter the dielectric constant of the solvent, which is in turn expected to change the affinity of metal ions to their cognate binding sites within an RNA molecule [418]. Molecular crowding conditions have been mimicked in other areas of research through addition of polyethylene glycol (PEG), which led to dramatic shifts in the structural equilibrium [419]. Future single-molecule studies are anticipated to be performed under such near-physiological conditions, which is likely to have an impact on the kinetic heterogeneity of nucleic acid samples.

## 6.5 Conclusion

The advent of single-molecule spectroscopy has provided a means to directly observe kinetic heterogeneities in biomolecular motion. Our recent report on heterogeneous docking/undocking behavior of d3'EBS1* and IBS1* takes its place among numerous other examples of complex kinetic behavior unveiled by single-molecule techniques. While we and others propose approaches to characterise molecular heterogeneity [387, 413], discord persists over its possible origins that are likely to be diverse. Recent findings indicate, however, that kinetic heterogeneities in *in vitro* single-molecule studies are not an experimental artefact, given that the experimental design is chosen with care and suitable control experiment are carried out. The real challenge will be to find out whether kinetic heterogeneity exists *in vivo*. We believe that recent efforts in developing *in vivo* single-molecule techniques will provide novel proofs of principle and further our understanding of this fascinating phenomenon.

# Chapter 7

# Cation-promoted RNA tertiary structure formation dissected by single-molecule fluorescence

Sebastian L.B. König,* Danny Kowerko,* Mokrane Khier, Roland K.O. Sigel

Dept. of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

* Equal contributions.

**ABSTRACT:** RNA folding and function are largely dependent on the action of cations that screen the negative charges along the phosphate-sugar backbone, mediate the formation of a precise three-dimensional fold, and possibly participate in catalysis. Here, we monitor the cation-dependent formation of the intron-exon recognition complex of a group II intron by single-molecule fluorescence in order to dissect the role of the cationic cofactor in RNA-RNA strand association. Rigorous analysis of the kinetics and thermodynamics reveals that the metal ions serve as cofactors for both tertiary contact formation formation and interstrand dissociation. Perfect agreements with the characteristic metal ion complex stabilities along the extended Irving-Williams series shows that RNA-RNA structure formation relies on non-specific charge screening, while the dissociation rate depends on the disruption of specific cation-phosphate bonds. RNA-RNA duplex formation can thus be explained through the intrinsic coordination chemistry of the metal ion involved and single-molecule FRET is validated to systematically study metal ion-mediated formation of nucleic acid structure.

**Keywords:** RNA, structure, metal ions, Irving-Williams Series, smFRET

*Equal contributions. Submitted to *Nat. Chemistry* in February 2014.

## 7.1 Introduction

Ribonucleic acids (RNAs) are involved in uncountable processes of cellular metabolism, including protein synthesis and its regulation [28]. However, before an RNA molecule achieves functionality, secondary structure motifs, *i.e.* duplexes, hairpins, bulges, internal loops and junctions, must form and subsequently assemble into a precise three-dimensional structure [23, 251]. The highly anionic nature of the phosphate-sugar backbone entails that structure formation is orchestrated through the action of metal ions [422]. These metal ions typically coordinate to the negatively charged phosphate moieties and to various exocyclic atoms located within the purine/pyrimidine nucleobases (Figure 7.1A) [2]. The affinity of divalent metal ions to these binding sites has been studied in isolation using small molecule mimics. Here, the stability of acetate, benzimidazole, and imidazole complexes follows the order $Ba^{2+} < Sr^{2+} < Ca^{2+} < Mg^{2+} < Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} << Cu^{2+} > Zn^{2+} \sim Cd^{2+}$, a trend that is commonly known as Irving-Williams series [420, 421, 423] (Figure 7.1B). In turn, the Irving-Williams series is broken when the metal ion coordinates to phosphate monoesters, which display a local maximum in stability in the presence of $Mn^{2+}$ (Figure 7.1B) [424].

Owing to their elevated bioavailability in mammalian cells, most of the studies addressing the influence of cations on structure and folding of RNAs precedented to date have focused on the role of $Na^{+}$, $K^{+}$, and $Mg^{2+}$, the latter of which is commonly believed to be the natural cofactor RNA folding and catalysis [65, 66]. Nonetheless, other cation species are known to reach

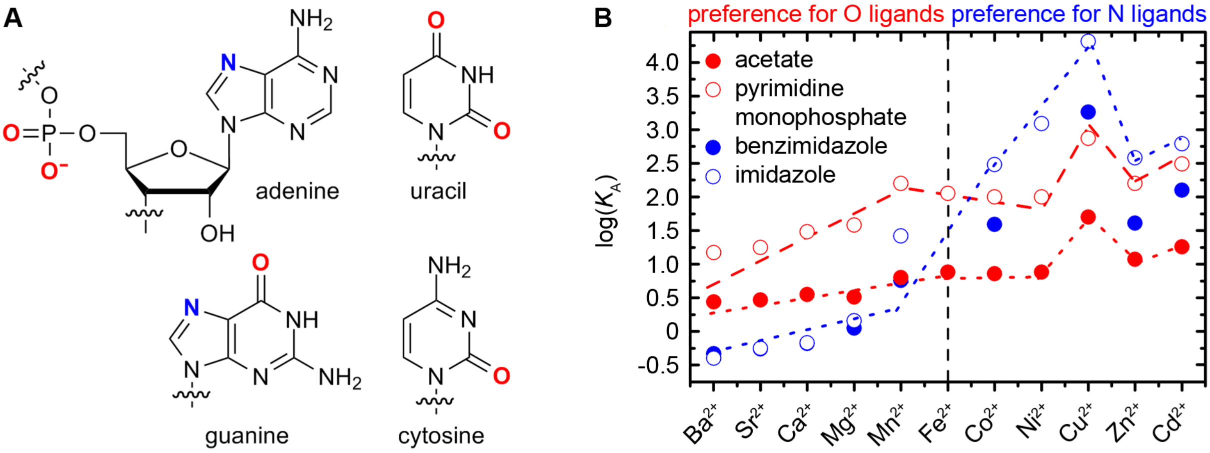


Figure 7.1: (A) Metal ion coordination to the RNA backbone and nucleobases. Depiction of the major cation binding sites occurring within the building blocks of RNA (coloured atoms). Figure modified from [107] (B) Stability of metal ion complexes using ligands mimicking metal ion binding sites within RNA. While the stability of acetate, benzimidazole, and imidazole complexes follow the Irving-Williams series, a local maximum is observed for manganese phosphate monoester complexes. A preference for oxygen ligands is observed for $Ba^{2+}$, $Sr^{2+}$, $Ca^{2+}$, $Mg^{2+}$, and $Mn^{2+}$, while $Co^{2+}$, $Ni^{2+}$, $Zn^{2+}$, and $Cd^{2+}$ preferably coordinate to oxygen. Lines represent a visual guide to the data. Data taken from [64, 107, 420, 421].

noteworthy concentrations inside cells and their compartments and may thus also be involved in RNA structure formation (see also: Section 3.1.6). For example, $Ca^{2+}$, which may reach intracellular concentrations in the low millimolar range, has recently been shown to induce misfolding of self-splicing group II introns *in vitro* [66, 78, 87, 102]. Other ribozymes display elevated reaction rates in the presence of rather exotic metal ions, such as the hammerhead ribozyme in the presence of $Mn^{2+}$, $Co^{2+}$, $Ni^{2+}$, $Zn^{2+}$, and $Cd^{2+}$ [91, 101, 107]. In these cases, $Mg^{2+}$ may not be the natural cofactor and/or ribozyme folding could be controlled via a metal ion switch. However, knowledge on metal ion selection by nucleic acids is still very scarce [107].

Based on high resolution structures, it is commonly believed that roughly 10 % of all cations interact with RNA in a highly specific manner, *i.e.* specific cation species bind to specific coordination sites [2, 107]. However, structural details of metal ion binding to complex nucleic acids are challenging to obtain and often fail to explain RNA conformational dynamics due to metal ions [238, 425]. Experimental approaches to characterise the dynamic interplay of metal ions with RNA and cation-mediated local structure formation are even more challenging and have to the best of our knowledge not been reported. Here, we employ single-molecule Förster resonance energy transfer (smFRET) to understand RNA-RNA tertiary structure formation and dynamics in dependence of divalent metal ions [109]. For this purpose, a previously validated sequence pair derived from the 5' splice site recognition complex in the group IIB intron Sc.ai5γ (d3'EBS1*/IBS1*) was labelled with the FRET pair Cy3 and Cy5. A biotin moiety was further attached to the 3'-end of d3'EBS1*. This allowed us to follow the interaction of single

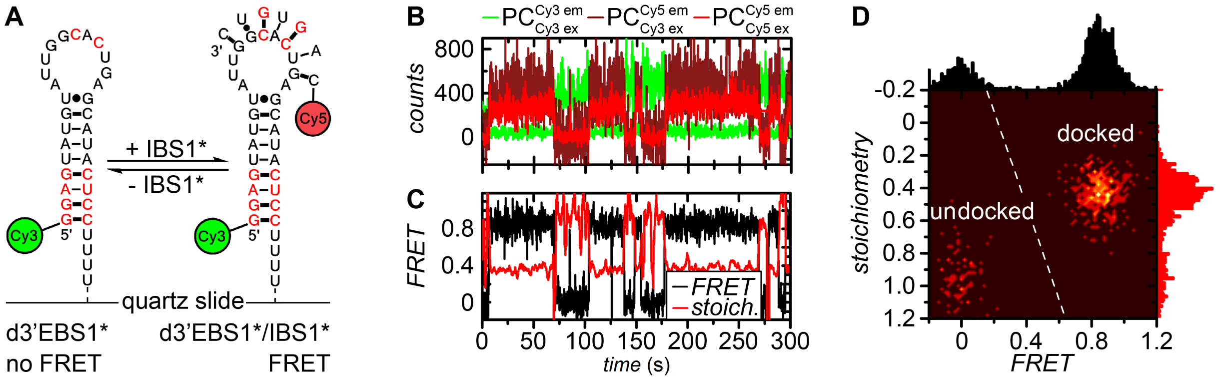


Figure 7.2: Experimental design and representative data. (A) Docking of IBS1* onto the surface-immobilised d3'EBS1* is accompanied by appearance of FRET from Cy3 to Cy5. (B) Representative fluorophore emission time trace. Cy3 (green) and Cy5 (red) display anticorrelated emission when Cy3 is excited at 532 nm. The presence of IBS1* is further detected by direct excitation of Cy5 (dark red). (C) *FRET* and *stoichiometry* over *time* as calculated from fluorophore emission time traces. The *stoichiometry* trace was smoothed by rolling-point averaging. (D) *FRET* - *stoichiometry* histogram binned from the data in shown in panel C reveals two clusters: The undocked state is characterised by the absence of both Cy5 emission and *FRET*, the docked state is characterised by the presence of both fluorophores in a 1:1 ratio and the occurrence of *FRET*.

surface-immobilised d3'EBS1* molecules with IBS1* in the presence of 9 different divalent cations along the Irving-Williams series (Figure 7.2). In contrast to bulk methods like X-ray diffraction or NMR that yield data on ensembles of billions of molecules, no averaging takes place when distance changes between a single FRET pair are recorded, hence unveiling entire folding pathways including scarcely populated intermediates [141, 161]. Our results show for the first time, that association and dissociation of the d3'EBS1*/IBS1* system follows two steps in each direction: While the rates of the rate determining association step follows the classical Irving-Williams series, dissociation follows phosphate coordination, dissecting the role of the metal ion cofactor in unprecedented detail.

## 7.2 Materials and Methods

### 7.2.1 Oligonucleotides and chemicals

Labelled RNA strands were ordered PAGE-purified from Microsynth AG (Balgach, Switzerland) and additionally HPLC purified [426]. All chemicals were purchased from Sigma-Aldrichs (Buchs, Switzerland).

### 7.2.2 UV thermal melting experiments

Temperature-dependent absorption profiles were measured at 260 nm using a Cary 500 Scan UV-vis-NIR spectrophotometer equipped with a Cary Temperature Controller (Varian inc., Palo Alto, CA, USA). RNA samples were dissolved in 50 mM MOPS, 100 mM $KNO_3$, 1 mM $M(NO_3)_2$ (optional), pH 6.90 to a consistent concentration of 3 µM per strand. Each sample

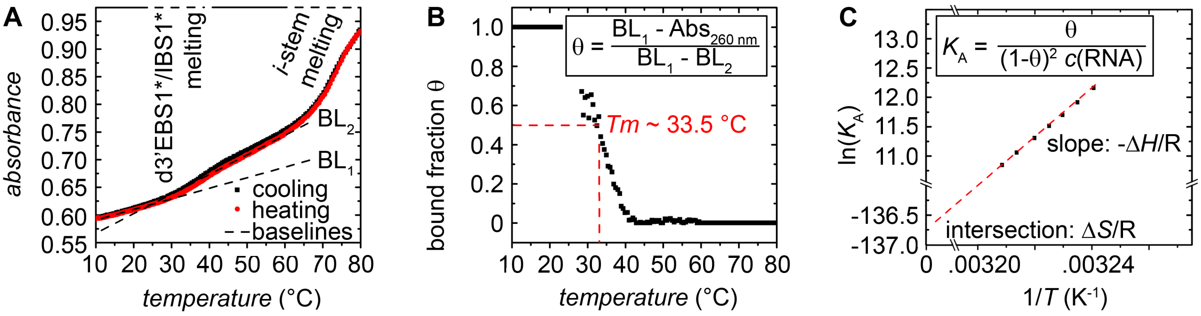


Figure 7.3: Thermal melting experiments and van't Hoff analysis, representative data. (A) Temperature-dependent *absorption* was followed at 260 nm during cooling and subsequent re-heating. Hysteresis is negligible at a temperature changing rate of 0.25 °C/min, as both curves superimpose well. Based on previous results, the first hyperchromic shift was assigned to d3'EBS1*/IBS1* dissociation, the second one to *i*-stem opening [388]. (B) Determination of the folded fraction $\theta$ was calculated from the baselines as precedented [389]. *Tm* is the temperature at which docked and undocked d3'EBS1* are at equimolar concentration. (C) Determination of thermodynamic parameters. The data shown were recorded in 50 mM MOPS, 100 mM $KNO_3$, 1 mM $Mn(NO_3)_2$, c(RNA) = 3 µM per strand, pH 6.9.

was degassed and overlaid with 400 µl mineral oil to avoid the formation of bubbles and evaporation. Measurements were carried out using Hellma 105.201-QS quartz cuvettes (Hellma GmbH & Co. KG, Müllheim, Germany) and a total number of three absorption ramps were recorded at a temperature changing rate of 0.25 °C $\text{min}^{-1}$ (85 °C - 10 °C - 85°). The sample chamber was flushed with $N_2$ during measurements to avoid condensation of $H_2O$ on the outer surface of the quartz slide at low temperature [427]. Measurements were repeated three times, followed by van't Hoff analysis to extract thermodynamic parameters (Figure 7.3) [17].

### 7.2.3 Single-molecule FRET measurements

Microfluidic chambers were prepared from quartz slides (Finkenbeiner, Waltham, MA, USA) as described [412]. The surface of the channel was passivated with biotinylated BSA, followed by tethering Cy3-labelled d3'EBS1* to the surface via a biotin-streptavidin linkage [170]. The imaging buffer contained 50 mM MOPS, 100 mM $KNO_3$, 1 mM $M(NO_3)_2$ (optional), 1 % D-glucose, 165 U/ml glucose oxidase, 2170 U/ml catalase, 1 mM Trolox, 25 nM Cy5-labeled IBS1*, pH 6.90. Cy3 and Cy5 emission profiles were monitored over 6 minutes at a time resolution of 100 ms using a home-built TIRF microscope upon ALEX as described (see also: Section 12.1) [354, 390].

### 7.2.4 smFRET imaging and data processing

Data analysis was performed using the Matlab-based software package SIRA, which has been described in detail elsewhere [390]. Clips were corrected for background noise and leakage of Cy3 emission into the Cy5 channel. Spot detection was followed by co-localisation of corresponding coordinates in the donor and acceptor channel. Photon counts (*PC*) were integrated over 3x3 pixels around the central coordinates. This yielded three time-dependent photon counts: Cy3 emission upon Cy3 excitation ($PC(t)^{\text{Cy3 em}}_{\text{Cy3 ex}}$), Cy5 emission upon Cy5 excitation ($PC(t)^{\text{Cy5 em}}_{\text{Cy5 ex}}$), and Cy5 emission upon Cy3 excitation ($PC(t)^{\text{Cy5 em}}_{\text{Cy3 ex}}$ , Figure 7.2B). Emission time traces were manually selected for anticorrelation to calculate time-dependent apparent *FRET* efficiencies and *fluorophore stoichiometries* (Figure 7.2C):

$$FRET(t) = \frac{PC(t)^{\text{Cy5 em}}_{\text{Cy3 ex}}}{PC(t)^{\text{Cy3 em}}_{\text{Cy3 ex}} + PC(t)^{\text{Cy5 em}}_{\text{Cy3 ex}}} \tag{7.1}$$

$$stoichiometry(t) = \frac{PC(t)^{\text{Cy5 em}}_{\text{Cy3 ex}} + PC(t)^{\text{Cy5 em}}_{\text{Cy5 ex}}}{PC(t)^{\text{Cy3 em}}_{\text{Cy3 ex}} + PC(t)^{\text{Cy5 em}}_{\text{Cy3 ex}} PC(t)^{\text{Cy5 em}}_{\text{Cy5 ex}}} \tag{7.2}$$

The resulting *FRET* and *fluorophore stoichiometry* time traces were then binned to 2D histograms (Figure 7.2D). 150 or more of such 2D histograms were summed up to determine the averaged bound fraction and the association constant $K_A$:

$$K_{\mathrm{A}} = \frac{c(\mathrm{d3'EBS^*/IBS1^*})}{c(\mathrm{d3'EBS^*})c(\mathrm{IBS1^*})} = \frac{p(FRET_{\mathrm{high}})}{p(FRET_{\mathrm{zero}}) + p(FRET_{\mathrm{high}})} \tag{7.3}$$

where $p(FRET_{\mathrm{high}})$ and $p(FRET_{\mathrm{zero}})$ correspond to the bound and unbound fraction determined by thresholding [413]. Kinetic analyses were performed using the dwell times, *i.e.* the time a molecule resides in the zero and the high *FRET* state. Dwell times were determined through thresholding, binned to histograms, plotted as normalised cumulative probabilities, and fitted to a stretched exponential decay model:

$$1 - cumP = \exp\left\{-\left(\frac{t}{\tau_{1/e}}\right)^{\beta}\right\} \tag{7.4}$$

where *cumP* is the normalised cumulative probability that a molecule dwells in the *FRET* state. *t* denotes the dwell time and the decay constant $\tau_{\mathrm{1/e}}$ refers the time required for 1 - *cumP* to drop to 1/e of its initial value [428]. The stretching exponent $\beta$ ($0 < \beta \leq 1$) is a means to quantify the distribution of rates, where 1 is equivalent to a single-exponential decay [428]. Subsequently, the rate constants for undocking ($k_{\mathrm{undocking}}$) and docking ($k_{\mathrm{docking}}$) were calculated from $\tau_{1/e}$ and $\beta$ :

$$k_{\mathrm{undocking}} = \frac{1}{\langle \tau_{\mathrm{1/e,\ high}} \rangle} = \left(\left(\frac{\tau_{\mathrm{1/e,\ high}}}{\beta_{\mathrm{high}}}\right)\Gamma\left(\frac{1}{\beta_{\mathrm{high}}}\right)\right)^{-1} \tag{7.5}$$

$$k_{\mathrm{docking}} = \frac{1}{\langle \tau_{\mathrm{1/e,\ high}} \rangle} = \left(\left(\frac{\tau_{\mathrm{1/e,\ high}}}{\beta_{\mathrm{high}}}\right)\Gamma\left(\frac{1}{\beta_{\mathrm{high}}}\right)c_{\mathrm{tot}}(\mathrm{d3'EBS1^*})\right)^{-1} \tag{7.6}$$

where $\Gamma$ the so-called gamma function [428].

### 7.2.5 Bootstrapping and statistical analysis

The variabilities associated with $K_{\mathrm{A}}$, $k_{\mathrm{docking}}$, and $k_{\mathrm{undocking}}$ were estimated using the freely available Matlab script BOBA FRET [166]. Briefly, histograms and dwell times from *N* individual time traces were resampled in order to build *M* bootstrapped cumulated *FRET* or dwell time histograms, respectively. Here, *N* corresponds to the number of individual time traces and *M* was set to 100 following the conventions from other fields. Subsequently, bootstrapped cumulated histograms were converted into probability density functions through normalisation and fitted to multiple Gaussian or stretched exponential decay models (Equations 7.6 and 7.7), respectively:

$$\theta(FRET) = \frac{1}{N}\sum_{i=1}^{m} A_i \exp\left\{-\frac{(FRET - b_i)^2}{2\sigma_i^2}\right\} \tag{7.7}$$

where *A* denotes the amplitude, $\sigma$ is the width and *b* is the center of the Gaussian distribution. *m* is the total number of Gaussian fit functions and was set to 2 as a consequence of the way the experiments were designed. Subsequently, the bootstrapped means and the corresponding standard deviations were calculated according to:

$$\sigma_{bootstrap} = \sqrt{\sum_{k=1}^{M} \frac{(X_k - \bar{X})^2}{M-1}} \tag{7.8}$$

Here, $X$ denotes the parameter whose variability is estimated. A detailled description of the mathematical groundwork and the software used Is given in Chapter 8 [166].

One-way analysis of variance (ANOVA) was performed in Matlab (version 8.20.701, license 49040, MathWorks, Nattick, MA, USA) using the statistics toolbox. Bonferroni's correction was used to correct for multiple testing (*ctype*) and the significance level (*alpha*) was set to 0.001.

### 7.2.6 Bulk fluorescence experiments

Bulk fluorescence experiments were carried in 96-well plates using a Safire 2 Microplate Reader (Tecan Group Ltd., Männedorf, Switzerland) adapted for rapid sequential determination of absorption and fluorescence spectra, as well as fixed wavelength measurements of anisotropy, fluorescence intensity, and absorbance. Coated 3995-COR 96-well plates (Vitaris, Baar, Switzerland) were used to prevent non-specific interaction of the sample and the surface. Measurements were conducted using sulfonated cyanine dye NHS esters (GE Healthcare, Glattbrugg, Switzerland) or sulfonated cyanine dyes tagged to RNA (d3'EBS1* or IBS1*), dissolved in imaging buffer at a consistent concentration of 100 nM. The imaging buffer contained 50 mM MOPS, 100 mM $KNO_3$, 1 % D-glucose, 165 U/mL glucose oxidase, 2170 U/mL catalase, 1 mM Trolox, pH 6.90. It further contained 0, 1, 10, 100 or 1000 mM $M(NO_3)_2$ where M may refer to Ba, Sr, Ca, Mg, Mn, Ni, Co or Cd. Each measurement was repeated three times to calculate the standard deviation.

#### Cation-induced Cy3/Cy5 fluorescence quenching

Fluorescence intensities were quantified using fixed wavelength scans at constant slit width ($\lambda_{\mathrm{exc}}^{\mathrm{Cy3}}$ = 530nm and $\lambda_{\mathrm{exc}}^{\mathrm{Cy5}}$ = 600nm). The emission intensity was determined at 560 and 660 nm, respectively. The exposure time was set to 80 µs and the gain was adjusted to the brightest well, followed by performing 2 reads per well.

#### Shape of Cy3/Cy5 fluorescence spectra in response to different metal ions

Fluorescence spectra were recorded three times, averaged and smoothed using a sliding average filter with a window of 10 data points. Finally, spectra were normalised to a maximum intensity of 1.

**Metal ion-dependent fluorescence anisotropy**

Fluorescence anisotropy $r$ was calculated as

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{7.9}$$

where $I_{\parallel}$ and $I_{\perp}$ denote the emission intensities when the emission polariser is oriented parallel or perpendicular to the direction of polarised excitation (see also: Section 3.2.5) [109].

## 7.3 Results

### 7.3.1 The thermodynamic stability of the d3'EBS1*/IBS1* complex is strongly affected by the identity of the divalent cation

Interaction of single Cy3-labelled d3'EBS1* molecules immobilised on the surface of a BSA-passivated quartz slide with Cy5-labelled IBS1* molecules in solution was monitored over several minutes at a temporal resolution of 100 ms using a TIRF microscope equipped with alternating laser excitation (ALEX) [170, 354]. Fluorophore emission over time was extracted from smFRET clips (Figure 7.2B), followed by calculating time-depedent apparent *FRET* and *fluorophore stoichiometry* (Equations 7.1 and 7.2, Figure 7.2C). This approach permits to exclude photophysical artefacts (blinking), the occurrence of doubly labelled molecules, several EBS1* molecules in irresolvable proximity and/or d3'EBS1*-d3'EBS1* duplexes from further analysis [354, 393]. Furthermore, the influence divalent metal ions at low millimolar concentrations was demonstrated to have a negligible effect on Cy3 and Cy5 fluorescence intensity and the shape of their fluorescence spectra, as well as their rotational diffusion (Figures 12.4, 12.5, 12.6, 12.7, 12.8 and 12.9 in the Appendix). *FRET* and *stoichiometry* trajectories were binned to 2D histograms (Figures 7.2D and 7.4). A considerable fraction of molecules was static, *i.e.* no *FRET* transitions were observed during the observation time (Figure 7.5). 15-55 % of Cy5 molecules are commonly known to pre-bleach before the measurement [146]. As free d3'EBS1* would not be distinguishable from docked d3'EBS1*/IBS1* bearing an inactive acceptor fluorophore, cumulated histograms were exclusively created from 2D plots built from dynamic time traces in order to avoid photophysical artefacts. Cumulated histograms were used to determine the bound fraction via integration (Figure 7.4), followed by calculating the association constant $K_A$ in dependence of the metal ion(s) present in the imaging buffer (Equation 7.3). Cross-sample variability associated with the bound fraction was estimated via bootstrapping using the freely available Matlab script BOBA FRET (Figure 7.6) [166].

In the absence of $M^{2+}$, the docked fraction approximates 7.8 % (Figure 7.4A), which corresponds to $\log K_A = 6.5$ (Figure 7.7). However, addition of 100 μM EDTA to the imaging buffer abolishes d3'EBS1*/IBS1* pairing almost completely and the vast majority of d3'EBS1* molecules remain statically undocked (Figure 7.4B). One can conclude that the $KNO_3$ salt used for

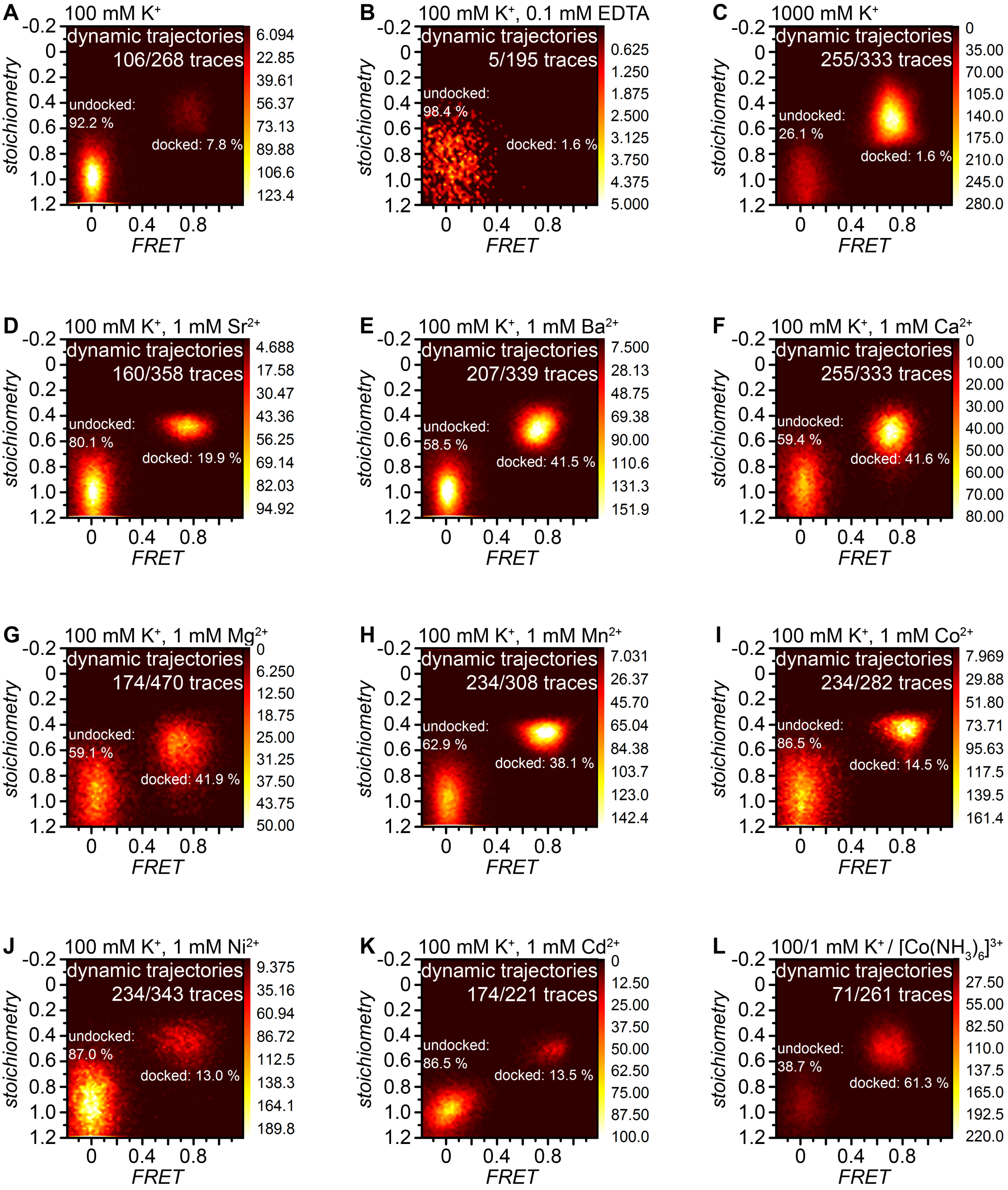


Figure 7.4: Normalised cumulated FRET - fluorophore stoichiometry histograms built from smFRET time traces recorded under various imaging conditions. Divalent metal ions strongly differ in their propensity to shift the thermodynamic equilibrium towards the docked conformation. The colour code refers to the occurrence in counts. Standard imaging conditions: 50 mM MOPS, 100 mM $KNO_3$, 25 mM Cy5-labeled IBS1*, 1 mM Trolox, 1x OSS, pH 6.9, additives as indicated.

the buffer preparation contains traces of divalent metal ions. Subsequently, we characterised d3'EBS1*/IBS1* interaction in the presence of different divalent cations along the extended Irving-Williams series (extIWS), each of which was added at 1 mM concentration. Experiments in the presence of $Fe^{2+}$ and $Cu^{2+}$ could not be evaluated due to instant degradation of the RNA, oxidation by $O_2$ and/or poor solubility at physiological pH [78]. $Zn^{2+}$ and $Cd^{2+}$ led to high levels of unspecific Cy5-IBS1* docking (Figure 7.8) as well as precipitation of $M(OH)_2$ in the case of $Zn^{2+}$. Consequently, quantification of these smFRET data proved difficult. However, a total number of 8 divalent metal ions could be assessed. In addition, the influence of high monovalent cation concentrations (1 M $K^+$) and the $[Mg(H_2O)_6]^{2+}$ mimic $[Co(NH_3)_6]^{3+}$ were also tested.

The presence of all divalent metal ions assessed herein induces a shift of the thermodynamic equilibrium towards the docked state, though the effect strongly varies in dependence of the identity of the cation (Figures 7.4, 7.6 and 7.7). In general, the averaged association constants follow the order $Ba^{2+} > Sr^{2+} < Ca^{2+} \sim Mg^{2+} \sim Mn^{2+} > Co^{2+} > Ni^{2+}$ (Figure 7.7). Analysis of variance (ANOVA) demonstrates the cation-induced changes are highly significant ($P <$ 0.001) [429]. Please refer to Table 7.1 for a summary of the *FRET* histogram analysis. Cation dependence of d3'EBS1*/IBS1* association constants was confirmed in UV thermal melting experiments, followed by van't Hoff analysis. Alterations in both $\Delta H^\circ$ and $\Delta S^\circ$ are observed in response to a change of the divalent cation, leading to $\log K_\mathrm{A}$ values that are generally in good agreement with the results of the smFRET experiments (Figure 7.7 and Table 7.2). Taken together, the thermodynamic stability of the d3'EBS1*/IBS1* complex is found to increase along the extended Irving-Williams Series, reaching a (local) maximum in the presence of $Ca^{2+}$, $Mg^{2+}$, and $Mn^{2+}$ (smFRET) or $Mn^{2+}$ (UV). Thermodynamic stability then traverses a

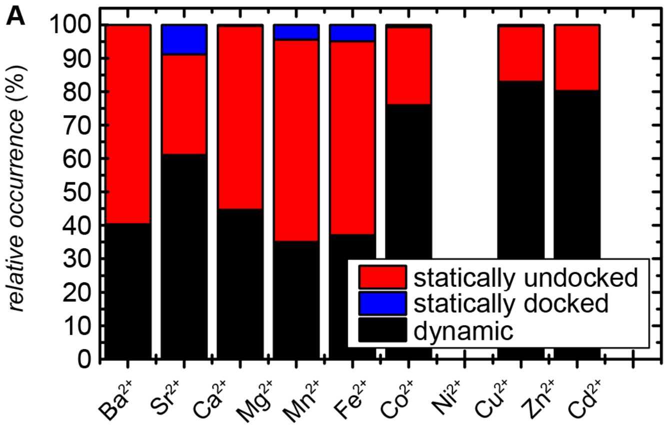


Figure 7.5: Classification of single-molecule *FRET* trajectories recorded in the presence of different divalent metal ions along the extended Irving-Williams series. Statically undocked: IBS1* does not dock during the observation time. Statically docked: IBS1* remains docked during the observation time. Dynamic: at least 1 docking or docking event during the observation time. Imaging buffer: 50 mM MOPS, 100 mM $KNO_3$, 25 mM IBS1*, 1 mM Trolox, 1x OSS, pH 6.9 and 1 mM $M(NO_3)_2$ as indicated.

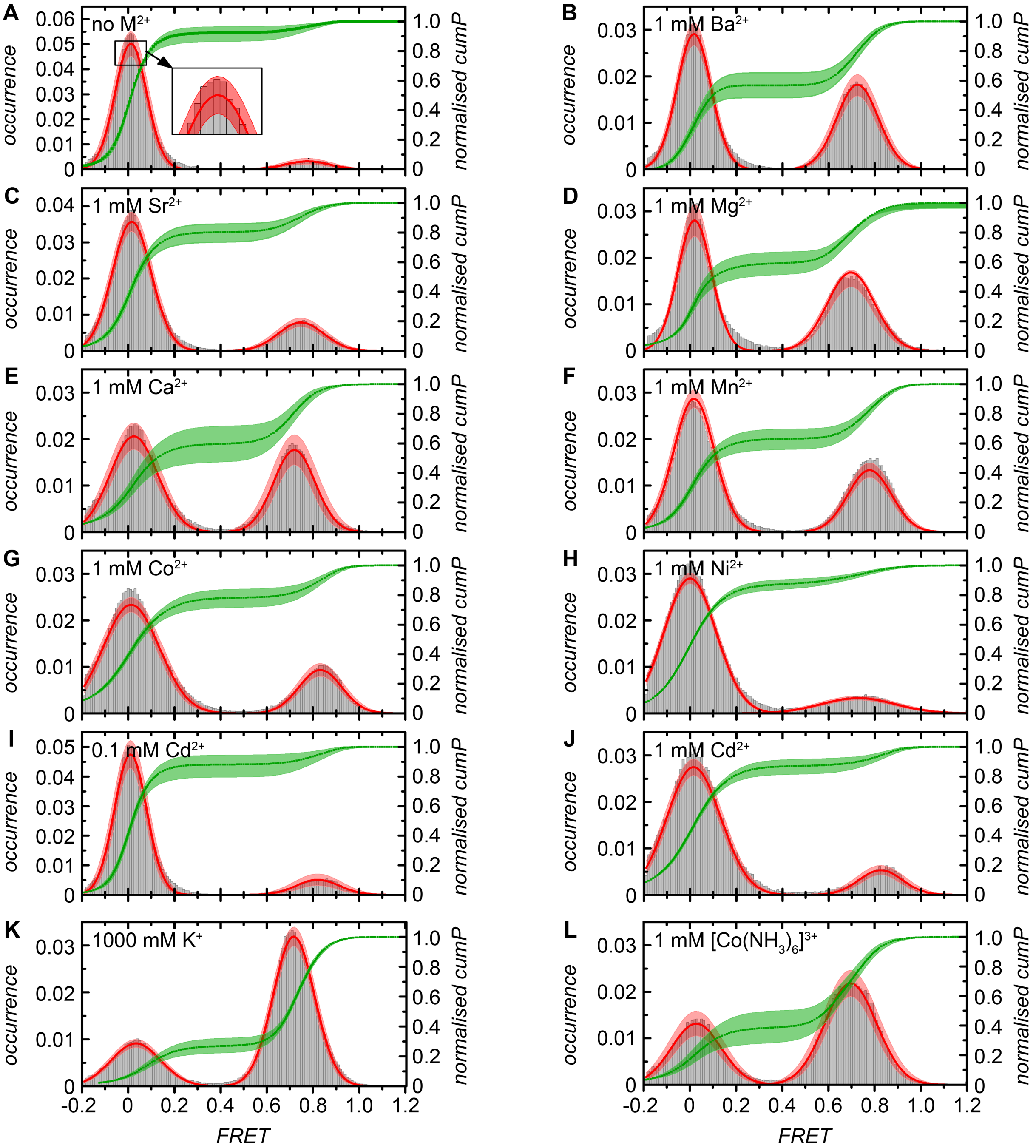


Figure 7.6: Multiple Gaussian fitting of cumulated *FRET* histograms and quantification of cross-sample variability via bootstrapping. Normalised *FRET* histograms were created from $\geq$71 *FRET* time trajectories and fitted to two Gaussian functions (zero: undocked, high: docked). The green curve corresponds to the normalised cumulative probability *cumP*. The red and yellow swaths correspond to the bootstrapped standard deviation ($2\sigma_{\text{bootstrap}}$) associated with the amplitude and the width of the Gaussian fit functions. Standard imaging buffer: 50 mM MOPS, 100 mM $KNO_3$, 25 mM IBS1*, 1 mM Trolox, 1x OSS, pH 6.9, additives as indicated.

minimum in the presence of $Co^{2+}$ or $Ni^{2+}$ (smFRET and UV), followed by a strong increase in interoligonucleotide affinity in response to the presence of $Zn^{2+}$ or $Cd^{2+}$ (UV).

### 7.3.2 Docking/undocking dynamics are greatly dependent on the identity of the divalent metal ion

Surface attachment of dye-labeled molecules allows to follow the dynamics of single molecules over several minutes, and thus, discern kinetic subpopulations [141, 161]. Differences in docking/undocking kinetics were quantified through dwell time analysis, *i.e.* binning dwell times in the zero and in the high *FRET* state to histograms, followed by approximation of the experimental data with a stretched exponential decay model (Equation 7.4 and Figure 7.9A-C). Hence, rate constants for IBS1* docking ($k_{\text{docking}}$) and undocking ($k_{\text{undocking}}$) could be determined for the presence of all cations assessed herein (Equations 7.5 and 7.6). Cross-sample variabilities associated with averaged decay constants and the respective stretching factors were estimated via bootstrapping (Figure 7.10) [166].

The docking rate is found to be 0.93 $\mu M^{-1}$ $s^{-1}$ in the presence of 100 mM $K^{+}$. In the presence of 1 mM divalent metal ions, docking rates steadily increase along the extended Irving-Williams series and they reach a maximum of 1.35 $\mu M^{-1}$ $s^{-1}$ in the presence of $Ni^{2+}$ (Figure 7.9D). However, it should be noted that relative fluctuations in $k_{\text{docking}}$ do not exceed a factor of 2. The rate associated with IBS1* undocking is determined to be 0.142 $s^{-1}$ in the absence of further additives. It is found to fluctuate around 0.05 $s^{-1}$ in the presence of $Ba^{2+}$, $Sr^{2+}$, $Ca^{2+}$ and $Mg^{2+}$. Subsequently, it reaches a minimum of 0.037 $s^{-1}$ in the presence of $Mn^{2+}$ (Figure 7.9D). IBS1* undocking is significantly accelerated upon addition of $Co^{2+}$ (0.112 $s^{-1}$)

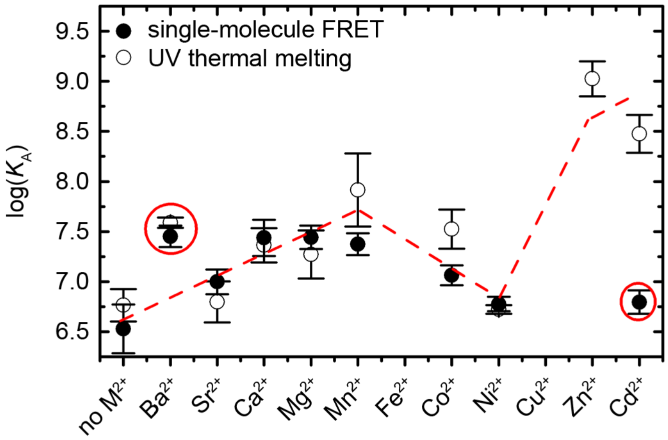


Figure 7.7: Thermodynamic stability of d3'EBS1*/IBS1* in the presence of different divalent cations as determined by single-molecule FRET and temperature-dependent UV absorption. The error bars associated with smFRET data were estimated from the bootstrapped values of the bound fraction ($3\sigma_{\text{bootstrap}}$), the experimental error associated with UV melting data were determined from replicates ($1\sigma$). Red circles indicated special points.

and $Ni^{2+}$ (0.189 $s^{-1}$), the latter of which leads to undocking rates higher than in the absence of divalent metal ions (Figure 7.9D). Fast IBS1* undocking is also observed in the presence of 1 mM $Cd^{2+}$ (0.116 $s^{-1}$), though, it is highly likely that undocking was biased by the surface passivation issues illustrated in Figure 7.8. Computing the rate constant from the thermodynamic stability determined by UV thermal melting and from the docking rate at 1 mM $Cd^{2+}$ leads to an estimated rate constant of 0.003 $s^{-1}$. In conclusion, both the values of $k_{\text{docking}}$ and $k_{\text{undocking}}$ respond to the presence of divalent metal ions, albeit the effect is considerably more pronounced for IBS1* undocking. The combined kinetic data are summarised in Table 7.3.

## 7.4 Discussion

Even though nucleic acid - metal ion interaction is a prerequisite for the formation of higher-order architectures, systematic studies on cation-dependent RNA folding are scarce, which is partly due to the limits of current experimental approaches [2, 430]. Here, smFRET was used to investigate the cation-mediated formation of the d3'EBS1*/IBS1* 5' splice site recogni-

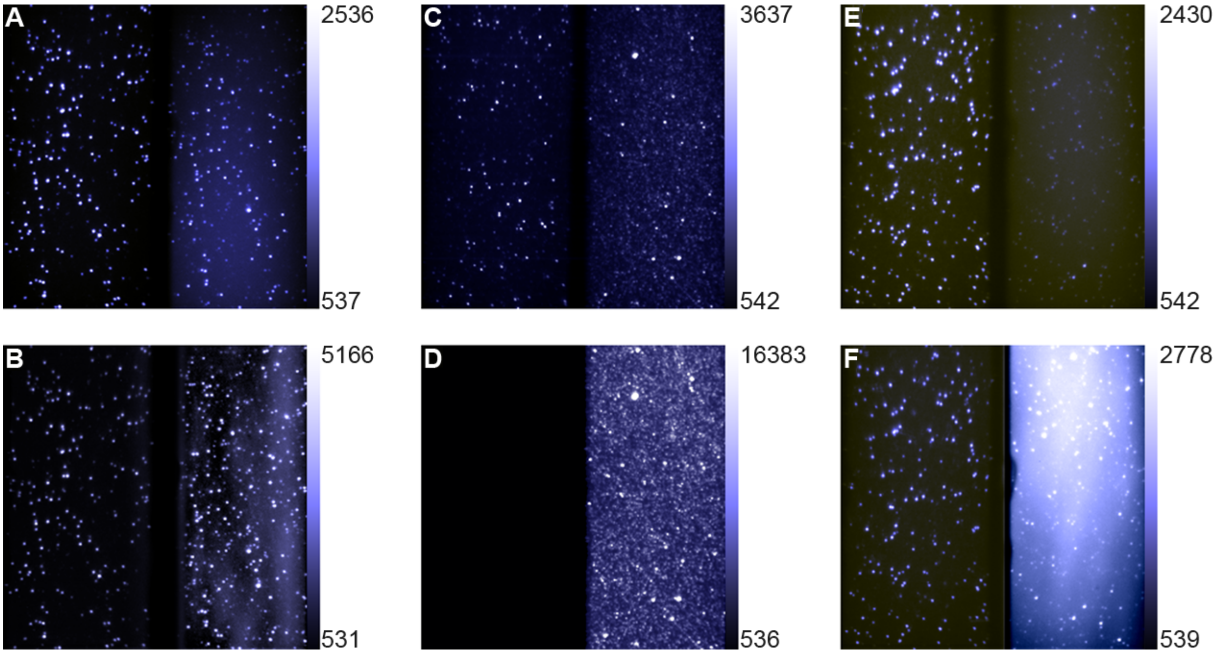


Figure 7.8: Non-specific docking of Cy5-labeled IBS1* at different imaging conditions. smFRET clips (Cy3 emission: left, Cy5 emission: right) were recorded using alternating laser excitation at 532 nm and 642 nm and averaged over frames recorded at 532 nm excitation (upper images) or over all frames (lower images). The colour code is in photon counts per $10^{-4}$ s per pixel (image resolution: 256 x 256 pixels). (A, B) Low levels of unspecific docking is observed in the presence of most divalent metal ions. Shown here: 1 mM $Mn^{2+}$. (C, D) Non-specific docking event frequently occur in the presence of 1 mM $Cd^{2+}$ as revealed when fluorophores are excited at 642 nm. (E, F) Elevated levels of background noise persist in the presence of 0.1 mM $Cd^{2+}$. Standard imaging buffer: 50 mM MOPS, 100 mM $KNO_3$, 25 mM IBS1*, 1x OSS, 1 mM Trolox, pH 6.9, divalent metal ions as described.

Table 7.1: Thermodynamic and kinetic parameters of d3'EBS1*/IBS1* interaction as determined by single-molecule fluorescence and UV spectroscopy. The experimental error is corresponds to $1\sigma$ or $1\sigma_{\mathrm{bootstrap}}$, respectively.

| Imaging conditions | High *FRET* centre (*FRET*) | High *FRET* width (*FRET*) | Docked fraction (%) | $\Delta G^{\circ}_{\mathrm{smFRET}}$ (kJ mol$^{-1}$) | $\Delta G^{\circ}_{\mathrm{UV}}$ (%) | Dynamic traces (%) | Static traces (%) |
|---|---|---|---|---|---|---|---|
| 100 mM K$^{+}$ | 0.776 ± 0.016 | 0.096 ± 0.005 | 7.8 ± 1.1 | -37.3 | -38.6 ± 0.9 | 40.3 | 59.7 |
| 1000 M K$^{+}$ | 0.717 0.004 | 0.090 0.003 | 73.9 ± 1.0 | -46.0 | N.D. | 76.6 | 23.4 |
| 100 mM K$^{+}$, 0.1 mM EDTA | N.D. | N.D. | 1.6 | -33.2 | N.D. | 2.5 | 97.5 |
| 100 mM K$^{+}$ 1 mM Ba$^{2+}$ | 0.728 ± 0.010 | 0.091 ± 0.004 | 41.5 ± 2.2 | -42.5 | -43.3 ± 0.3 | 61.1 | 38.9 |
| 100 mM K$^{+}$, 1 mM Sr$^{2+}$ | 0.748 ± 0.010 | 0.098 ± 0.004 | 19.9 ± 1.7 | -39.9 | -38.8 ± 1.2 | 44.7 | 55.3 |
| 100 mM K$^{+}$, 1 mM Ca$^{2+}$ | 0.719 ± 0.008 | 0.090 ± 0.005 | 40.6 ± 1.6 | -42.4 | -42.0 ± 2.1 | 35.1 | 64.9 |
| 100 mM K$^{+}$, 1 mM Mg$^{2+}$ | 0.697 ± 0.007 | 0.103 ± 0.003 | 40.9 ± 2.4 | -42.5 | -41.5 ± 1.4 | 37.0 | 63.0 |
| 100 mM K$^{+}$, 1 mM Mn$^{2+}$ | 0.779 ± 0.006 | 0.094 ± 0.003 | 37.1 ± 1.8 | -42.1 | -45.2 ± 2.1 | 76.0 | 24.0 |
| 100 mM K$^{+}$, 1 mM Co$^{2+}$ | 0.832 ± 0.006 | 0.092 ± 0.003 | 22.4 ± 1.5 | -40.3 | -42.9 ± 1.1 | 83.0 | 17.0 |
| 100 mM K$^{+}$, 1 mM Ni$^{2+}$ | 0.722 ± 0.013 | 0.160 ± 0.006 | 13.0 ± 0.8 | -38.7 | -38.4 ± 0.3 | 80.2 | 19.8 |
| 100 mM K$^{+}$, 1 mM Zn$^{2+}$ | N.D. | N.D. | N.D. | N.D. | -51.5 ± 1.0 | N.D. | N.D. |
| 100 mM K$^{+}$, 0.1 mM Cd$^{2+}$ | 0.821 ± 0.016 | 0.093 ± 0.006 | 12.0 ± 2.1 | -38.5 | N.D. | 45.8 | 54.2 |
| 100 mM K$^{+}$, 1 mM Cd$^{2+}$ | 0.825 ± 0.009 | 0.093 ± 0.006 | 13.5 ± 1.2 | -38.5 | -48.4 ± 1.0 | 82.0 | 18.0 |
| 100 mM K$^{+}$, 1 mM [Co(NH$_3$)$_6$]$^{3+}$ | 0.696 ± 0.012 | 0.112 ± 0.005 | 61.3 ± 3.5 | -44.5 | N.D. | 27.5 | 72.5 |

tion complex from group II introns. The results demonstrate that near-physiological potassium concentration does not suffice to ensure interoligonucleotide interaction to a significant extent (Figure 7.7). In turn, addition of 1 mM divalent metal ions significantly shifts the thermodynamic equilibrium towards the bound state, though, the effect strongly depends on the identity of the cation (Figure 7.7). The divalent metal ions that were assessed herein are known to display characteristic trends in the affinity towards specific metal ion binding sites located within nucleic acids (Figure 7.1) [424]. While the stabilities of acetate, imidazole and benzimidazole complexes follow the Irving-Williams series, it is a long-standing fact that this trend is broken for the stability of pyrimidine monophosphate complexes [64, 107, 420, 421, 424].

Table 7.2: Thermodynamic parameters of d3'EBS1*/IBS1* interaction at different buffer conditions as determined by van't Hoff analysis of thermal melting curves. Temperature-dependent absorption profiles were recorded in 50 mM MOPS, 100 mM $KNO_3$, 3μM RNA/strand, pH 6.90 and additives as indicated. The experimental error corresponds to the standard deviation $1\sigma$.

| Imaging conditions | Melting temperature (°C) | $\Delta H^\circ$ (kJ mol-1) | $\Delta S^\circ$ (J $K^{-1}$ $mol^{-1}$) | $\Delta G^\circ_{UV}$ (kJ $mol^{-1}$) | $\log(K_A)$ |
|---|---|---|---|---|---|
| 100 mM $K^+$ | 30.4 ± 0.9 | -304.0 ± 6.2 | -890.1 ± 18.0 | -38.6 ± 0.9 | 6.76 ± 0.16 |
| 100 mM $K^+$, 1 mM $Ba^{2+}$ | 33.3 ± 0.5 | -373.7 ± 6.2 | -1107.7 ± 21.9 | -43.3 ± 0.3 | 7.59 ± 0.05 |
| 100 mM $K^+$, 1 mM $Sr^{2+}$ | 30.5 ± 1.1 | -311.8 ± 3.0 | -915.2 ± 6.3 | -38.8 ± 1.2 | 6.80 ± 0.20 |
| 100 mM $K^+$, 1 mM $Ca^{2+}$ | 32.0 ± 0.4 | -385.1 ± 26.6 | -1150.4 ± 86.3 | -42.0 ± 1.0 | 7.36 ± 0.17 |
| 100 mM $K^+$, 1 mM $Mg^{2+}$ | 31.6 ± 1.2 | -385.0 ± 11.5 | -1151.6 ± 39.2 | -41.5 ± 1.4 | 7.27 ± 0.24 |
| 100 mM $K^+$, 1 mM $Mn^{2+}$ | 34.8 ± 1.8 | -378.0 ± 2.1 | -1115.9 ± 13.9 | -45.2 ± 2.1 | 7.92 ± 0.37 |
| 100 mM $K^+$, 1 mM $Co^{2+}$ | 33.7 ± 0.8 | -345.8 ± 7.2 | -1015.0 ± 20.3 | -42.9 ± 1.1 | 7.52 ± 0.19 |
| 100 mM $K^+$, 1 mM $Ni^{2+}$ | 30.5 ± 1.5 | -301.2 ± 9.8 | -881.0 ± 33.6 | -38.4 ± 0.3 | 6.72 ± 0.04 |
| 100 mM $K^+$, 1 mM $Zn^{2+}$ | 46.8 | -269.5 | -730.9 | -51.5 | 9.02 |
| 100 mM $K^+$, 1 mM $Cd^{2+}$ | 43.7 ± 1.3 | -256.7 ± 6.5 | -698.4 ± 20.4 | -48.4 ± 1.1 | 8.47 ± 0.19 |

Interestingly, the thermodynamic stabilities of the d3'EBS1*/IBS1* complex recorded in the presence of various different divalent metal ions display the phosphate-typical order: log$K_A$ values generally increase from $Sr^{2+}$ to $Mn^{2+}$, experience a decrease in the presence of $Co^{2+}$ and $Ni^{2+}$, and finally reach a maximum upon addition of $Zn^{2+}$ and $Cd^{2+}$ (Figure 7.7). Taken together, the combined results from UV melting and smFRET experiments suggest that the thermodynamic stability of the docked complex is largely governed by the intrinsic affinity of the cation towards phosphate moieties, which in turn means that phosphate residues must constitute the prime metal ion interaction site in the docked state. This is not surprising, as the nitrogen bases involved in canonical base pairing are generally sequestered in the interior of the duplex regions, while the phosphate backbone faces outward and is thus readily accessible for cation binding to take place [257]. Moreover, a recent NMR study suggests the presence of several specific $M^{2+}$ binding sites in d3'EBS1*/IBS1*, one of which is located at a strong kink in the d3'EBS1* backbone formed upon IBS1* docking [50]. Such arrangement of the RNA backbone brings several phosphate groups very close in space, thus explaining the requirement for divalent metal ions for stable interstrand association [50]. The importance of cations for the interaction of d3'EBS1* and IBS1* is also reflected by pronounced changes in the entropic contribution to $\Delta G°$ (Table 7.3) [17]. High interoligonucleotide affinities are observed in the

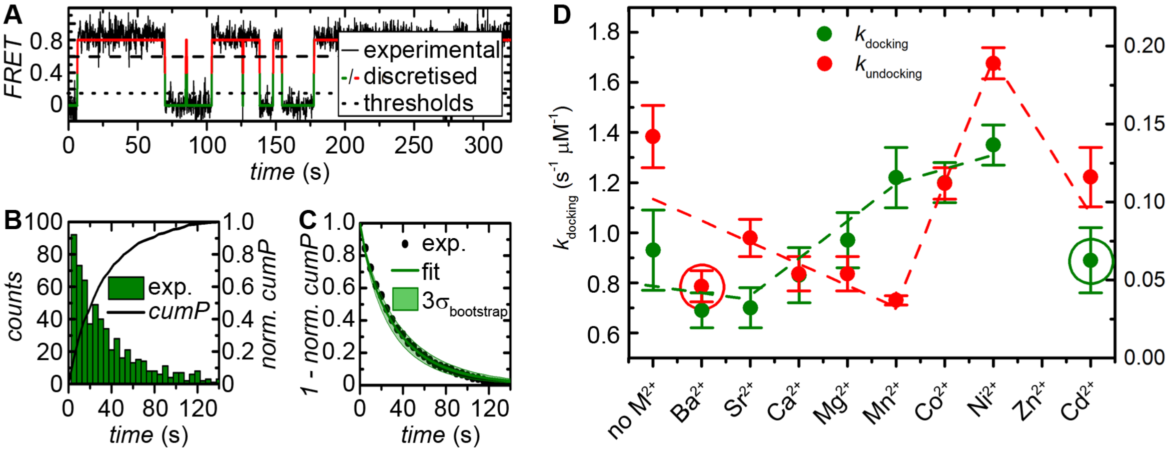


Figure 7.9: Dwell time analysis, representative data. (A) Discretised *FRET* trajectories are obtained through thresholding, which permits to extract dwell times in the zero (green) and high (red) *FRET* state. (B) Dwell times obtained from ≥ 150 time traces are binned to a histogram. The cumulative probability *cumP* corresponds to the normalised cumulative integral over the histogram. Data shown here are dwell times in the zero *FRET* state recorded at 100 mM $KNO_3$, 1 mM $Ni(NO_3)_2$. (C) 1 - *normalised cumP* is plotted over *time* and fitted to a stretched exponential decay model to obtain the rate constant *k* associated with docking or undocking. (D) Kinetics of IBS1* docking and undocking in the presence of different divalent cations. Green: Docking rates $k_{docking}$ as determined from dwell times in the zero *FRET*. Red: Undocking rates $k_{undocking}$ as determined from dwell times in the high *FRET*. Error bars were estimated by bootstrapping (1σbootstrap) [166]. Lines represent a visual guide to the data, while special points are encircled. Standard imaging conditions: 50 mM MOPS, 100 mM $KNO_3$, 25 mM IBS1*, 1x OSS, 1 mM Trolox, pH 6.9 and additives as indicated.

presence of $Ba^{2+}$, an observation that was attributed to differences in the preferred coordination geometry of the metal ion that is likely to lead to alternative metal ion-RNA contacts as reported for high-resolution structures [430].

Single-molecule spectroscopy provides a handle to resolve structural and kinetic heterogeneities hidden within the ensemble population [141, 161]. Quantification of cation-dependent docking

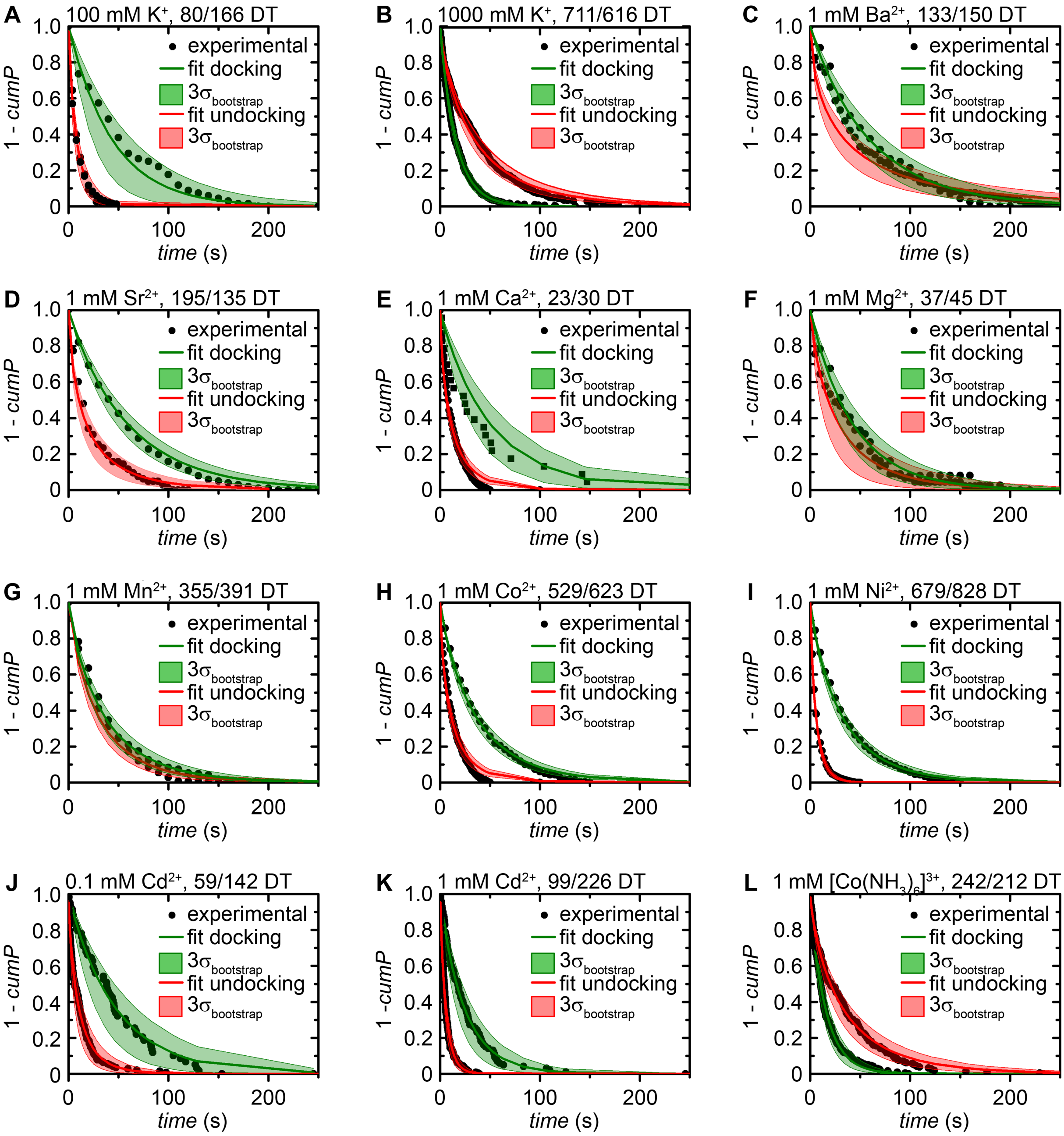


Figure 7.10: Dwell time analysis of IBS1* docking and undocking in the presence of divalent cations along the extended Irving-Williams Series. Standard imaging buffer: 50 mM MOPS, 100 mM $KNO_3$, 25 mM IBS1*, 1x OSS, 1 mM Trolox, pH 6.9 and an additives/modifications as indicated. The swaths correspond to the bootstrapped standard deviation $\sigma_{\text{bootstrap}}$ [166].

Table 7.3: Kinetic parameters of d3'EBS1*/IBS1* interaction at different buffer conditions as determined by single-molecule fluorescence. smFRET clips of surface-immobilised d3'EBS1* molecules were recorded in 50 mM MOPS, 100 mM KNO$_3$, 25 nM IBS1*, 1x OSS, 1 mM Trolox, pH 6.90 and additives as indicated. The experimental error corresponds to the estimated standard deviation $1\sigma_{\text{bootstrap}}$.

| Imaging conditions | $\tau_{1/e,\text{ zero}}$ (s) | $\beta_{\text{zero}}$ | $k_{\text{docking}}$ ($\mu$M$^{-1}$ s$^{-1}$) | $\tau_{1/e,\text{ high}}$ (s) | $\beta_{\text{high}}$ | $k_{\text{undocking}}$ s$^{-1}$ |
|---|---|---|---|---|---|---|
| 100 mM K$^+$ | 43.89 ± 6.72 | ~1.00 | 0.93 ± 0.16 | 7.97 ± 0.81 | 0.83 ± 0.07 | 0.142 ± 0.020 |
| 1000 mM K$^+$ | 15.56 ± 0.71 | ~1.00 | 2.60 ± 0.12 | 34.70 ± 2.10 | 0.79 ± 0.04 | 0.033 ± 0.003 |
| 100 mM K$^+$, 0.1 mM EDTA | - | - | - | - | - | - |
| 100 mM K$^+$, 1 mM Ba$^{2+}$ | 59.71 ± 5.33 | 0.96 ± 0.04 | 0.69 ± 0.07 | 35.05 ± 4.49 | 0.59 ± 0.05 | 0.046 ± 0.010 |
| 100 mM K$^+$, 1 mM Sr$^{2+}$ | 59.05 ± 5.70 | 0.96 ± 0.04 | 0.70 ± 0.08 | 17.67 ± 2.06 | 0.68 ± 0.05 | 0.077 ± 0.012 |
| 100 mM K$^+$, 1 mM Ca$^{2+}$ | 50.36 ± 6.32 | 0.95 ± 0.06 | 0.83 ± 0.11 | 52.47 ± 22.20 | 0.71 ± 0.05 | 0.054 ± 0.011 |
| 100 mM K$^+$, 1 mM Mg$^{2+}$ | 42.49 ± 6.67 | 0.95 ± 0.05 | 0.97 ± 0.11 | 28.36 ± 6.72 | 0.78 ± 0.05 | 0.044 ± 0.011 |
| 100 mM K$^+$, 1 mM Mn$^{2+}$ | 35.80 ± 2.53 | 0.87 ± 0.06 | 1.22 ± 0.12 | 29.66 ± 1.99 | 0.84 ± 0.04 | 0.037 ± 0.003 |
| 100 mM K$^+$, 1 mM Co$^{2+}$ | 35.57 ± 1.72 | 0.87 ± 0.04 | 1.20 ± 0.08 | 11.05 ± 0.68 | 0.72 ± 0.03 | 0.112 ± 0.010 |
| 100 mM K$^+$, 1 mM Ni$^{2+}$ | 31.80 ± 1.61 | 0.87 ± 0.04 | 1.35 ± 0.08 | 5.79 ± 0.23 | 0.84 ± 0.03 | 0.189 ± 0.010 |
| 100 mM K$^+$, 1 mM Zn$^{2+}$ | - | - | - | - | - | - |
| 100 mM K$^+$, 0.1 mM Cd$^{2+}$ | 46.98 ± 6.69 | 0.96 ± 0.04 | 0.89 ± 0.14 | 10.34 ± 1.35 | 0.77 ± 0.05 | 0.116 ± 0.019 |
| 100 mM K$^+$, 1 mM Cd$^{2+}$ | 24.20 ± 2.78 | 0.94 ± 0.05 | 0.89 ± 0.13 | 5.65 ± 0.43 | 0.91 ± 0.06 | 0.116 ± 0.019 |
| 100 mM K$^+$, 1 mM [Co(NH$_3$)$_6$]$^{3+}$ | 15.11 ± 1.18 | 0.96 ± 0.04 | 2.72 ± 0.23 | 32.24 ± 2.88 | 0.79 ± 0.05 | 0.036 ± 0.005 |

and undocking kinetics shows that docking rates steadily increase along the extended Irving-Williams Series, a trend that coincides strikingly well with the intrinsic cation affinities towards exo- and endocyclic heteroatoms located within ribonucleobases (Figures 7.1 and 7.9). As a consequence, IBS1* does most likely not involve the formation of (transient) rate-limiting metal ion - phosphate bonds. Instead, the role of the metal ion in the process of IBS1* docking is likely to be non-specific charge screening, as (i) cation-mediated increases of docking rate do not exceed a factor of 2 and (ii) docking could also be accelerated to a greater extent by adding large amounts of monovalent cations or 1 mM of the trivalent $[Mg(H_2O)_6]^{2+}$ mimic $[Co(NH_3)_6]^{3+}$ (Table 7.3) [405].

On the other hand, undocking rates generally decrease along the extended Irving Williams series, reaching a minimum in the presence of $Mn^{2+}$, followed by a strong decrease in the presence of $Co^{2+}$ and $Ni^{2+}$, which apparently fail to mediate stable d3'EBS1*/IBS1* interaction (Figure 7.9). Addition of $Cd^{2+}$ is estimated to cause a strong decrease of $k_{\text{undocking}}$. Hence, there is a pronounced influence of the cationic cofactor on undocking rates, which span more than one order of magnitude. The observed trend is in excellent agreement with the phosphate-

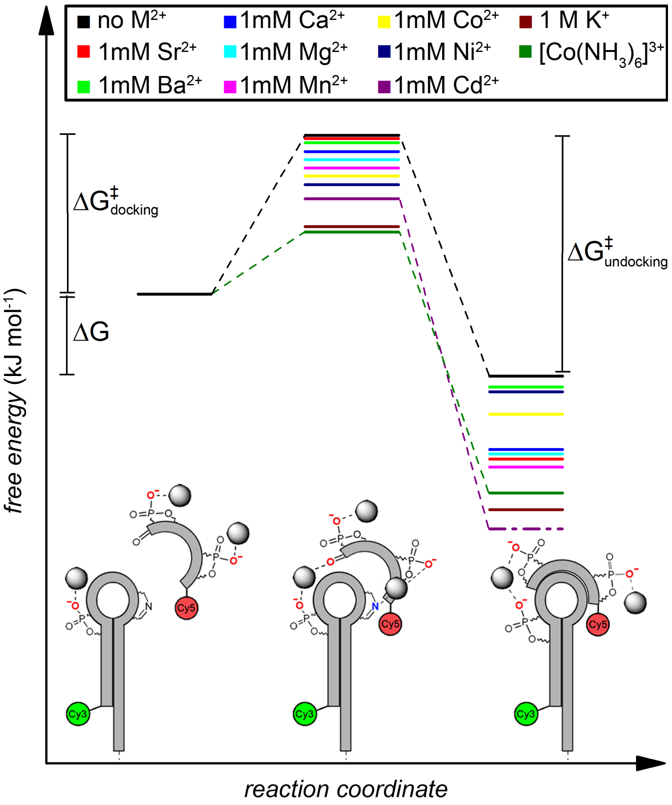


Figure 7.11: Working model for cation-mediated interaction of d3'EBS1* and IBS1* (normalised to the energy of the undocked state). In the undocked state, phosphate moieties are the prime cation binding site. d3'EBS1/IBS1* docking is non-specifically facilitated through charge screening. The docked state is specifically stabilised by coordination of the metal ions to phosphate residues.

affinities of the applied metal ions (Figure 7.1), demonstrating that the cation-induced shift of the thermodynamic equilibrium can for the most part be attributed to stabilisation of the d3'EBS1*/IBS1* complex and not an acceleration of the docking reaction. This in turn suggests that undocking most likely involves breaking specific metal ion phosphate bond mediating stable interstrand association.

## 7.5 Conclusion

The combined thermodynamic and kinetic data suggest the following mechanism for cation-mediated interaction of d3'EBS1* and IBS1* (Figure 7.11): Initially, metal ions are loosely and non-specifically bound to the phosphate backbone of the oligonucleotides because of electrostatic attraction [431]. When both RNA fragments come sufficiently close in space, cations form transient non-specific contacts with the nitrogen and oxygen moieties residing within the ribonucleobases and the riboses. This represents the rate limiting step in association of the two oligonucleotides as replacement of the metal ion leads to changes in the docking rate according to its intrinsic affinity towards oxygen and/or nitrogen ligands. These transient contacts are subsequently broken as the bases are sequestered within the double helix, prompting the metal ions to form the native phosphate contacts that mediate the formation of a stable interstrand architecture. Subsequent undocking of the two RNA fragments requires breaking of at least one phosphate-cation bond. The rate at which this second rate-limiting step occurs seems to depend almost exclusively on the affinity of the metal ion towards phosphate. Other factors such as the ionic radius, the preferred coordination number, the binding geometry, or the $\mathrm{p}K_A$ appear to be negligible (except for $\mathrm{Ba}^{2+}$).

The present studies validates smFRET as a robust tool for elucidating metal ion dependent folding of RNA using an unprecedented experimental design and rigorous statistical analysis. The findings show that RNA structure formation can be modulated by the intrinsic coordination chemistry of the metal ion cofactor, thus bridging the field of RNA biochemistry and coordination chemistry. These findings are along the lines of a recent study addressing the catalytic activity of the hammerhead ribozyme in the presence of various different metal ions whose mysteries could recently be solved by explaining cleavage rate with the phosphate affinity of the applied metal ion [91, 107]. As a consequence, the nature of the cationic cofactor appears to play a crucial role in both structure formation and catalytic activity. The present results are anticipated to prove valuable in understanding the principles underlying metal-ion mediated folding of nucleic acids and pave the way for studying larger and catalytically active ribozymes with smFRET.

# Chapter 8

# BOBA FRET: Bootstrap-based analysis of single-molecule FRET data

Sebastian L.B. König,[1] Mélodie Hadzic,[1] Erica Fiorini,[1] Richard Börner,[1] Danny Kowerko,[1] Wolf U. Blanckenhorn,[2] Roland K.O. Sigel[1]

1 Institute of Inorganic Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; 2 Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

**ABSTRACT:** Time-binned smFRET experiments with surface-tethered nucleic acids or proteins permit to follow folding and catalysis of single molecules in real-time. Due to the intrinsically low signal-to-noise ratio (SNR) in smFRET time traces, research over the past years has focused on the development of new methods to extract discrete states from noisy data. However, limited observation time typically leads to pronounced cross-sample variability, *i.e.* single molecules display differences in the relative population of states and the corresponding conversion rates. Quantification of cross-sample variability is necessary to perform statistical testing in order to assess whether changes observed in response to an experimental parameter (metal ion concentration, the presence of a ligand) are significant. However, such hypothesis testing has been disregarded to date, precluding robust biological interpretation. Here, we address this problem by a bootstrap-based approach to estimate the experimental variability. Simulated time traces are presented to assess the robustness of the algorithm in conjunction with approaches commonly used in thermodynamic and kinetic analysis of time-binned smFRET data. Furthermore, a pair of functionally important sequences derived from the self-cleaving group II intron *Sc*.ai5$\gamma$ (d3'EBS1*/IBS1*) is used as a model system. Through statistical hypothesis testing, divalent metal ions are shown to have a statistically significant effect on both thermodynamic and kinetic aspects of their interaction.

Published in *PLoS ONE* 2013, **8**, e84157.

## 8.1 Introduction

Förster Resonance Energy Transfer (FRET), distance-dependent energy transfer via a long-range dipole-dipole interaction, occurs between a donor fluorophore and an acceptor, which is typically (but not necessarily) also a fluorophore [109]. FRET results in a decrease in both donor emission intensity and lifetime, as well as the appearance of acceptor fluorescence [121]. Monitoring FRET between a single pair of dyes (smFRET) attached to a biomolecule can resolve both static and dynamic heterogeneity within a sample, *i.e.* differences between molecules and time-dependent conformational changes of individual molecules, both of which would otherwise be hidden through ensemble averaging [28, 189]. smFRET experiments are performed either on freely diffusing or upon surface attached molecules, the latter allowing for observation over an extended period of time. Technically, experiments with diffusing samples are implemented using a confocal microscope suitable for single-photon detection (time-correlated single photon counting, TCSPC). Experiments involving surface-tethered molecules can also be conducted with the aforementioned confocal microscope setup [432], although a wide-field or total internal reflection geometry is typically used for excitation, followed by detection with a CCD camera, resulting in time-binned *FRET* trajectories [141, 433]. Statistical analysis of such time-binned data is the objective of this article.

As smFRET data are generated from the emission of single fluorophores, the signal-to-noise ratio (SNR) is generally an issue, and considerable effort has been geared towards the development of tools to analyse noisy time traces. Ideally, such tools should permit to determine the number of conformational states in the system, their relative occurrence, and the rates at which they interconvert [192]. Cumulated *FRET* histograms have proven useful for simple two- or three-state systems, in which the approximation of individual *FRET* distributions with a normal distribution leads to minimal discrepancies [121]. When there is no or minimal overlap between the *FRET* distributions, the relative occurrence of the states is quantified by defining arbitrary cutoff values between *FRET* distributions (thresholding, Figure 8.1) [434]. In the case of moderate overlap, multiple Gaussian fits are typically performed to extract quantitative information (Figure 8.1) [435]. Under these circumstances, dwell times, *i.e.* the time spent in a certain *FRET* state until a conformational change occurs, can also be easily determined by thresholding, typically followed by fitting the dwell time histograms to exponential decay models to extract the rates of conformational rearrangement (Figure 8.1) [195, 196, 197, 219]. However, as the SNR deteriorates (short exposure times or fluorescence quenching ) or the centers of FRET distributions come close (similar interdye distances or modest conformational dynamics), these straightforward approaches can no longer be sensibly applied (Rayleigh criterion, Figure 8.1).

Noise in *FRET* time traces can be reduced through smoothing, *i.e.* by averaging out the inherent noise of the data collection process and hence emphasising the discrete nature of the *FRET* levels [284]. While linear rolling point averaging (also: moving or sliding averaging) is known to obscure transitions with dwell times shorter than the averaging window, the more sophisticated non-linear forward backward filter initially proposed by Chung and Kennedy and adapted by Haran partly overcomes this problem [391, 392]. Nevertheless, it also tends to average out very brief excursions to conformational intermediates in our hands. Taylor *et al.* recently presented an implementation of wavelet shrinkage to denoise *FRET* trajectories (Figure 8.1) [440, 441]. Here, the observed time series are transformed into a frequency component, followed by suppression of the noise assumed to lie within the high-frequency region of the transformation and inversion of the transformation that yields (in theory) a denoised dataset [440, 443]. It should be noted, however, that noise and signal often overlap in smFRET data, and thus such transformations may lead to spurious oscillations close to the transition (Gibb's phenomenon) [444]. A further application of wavelet transformation is termed change-point identification and has

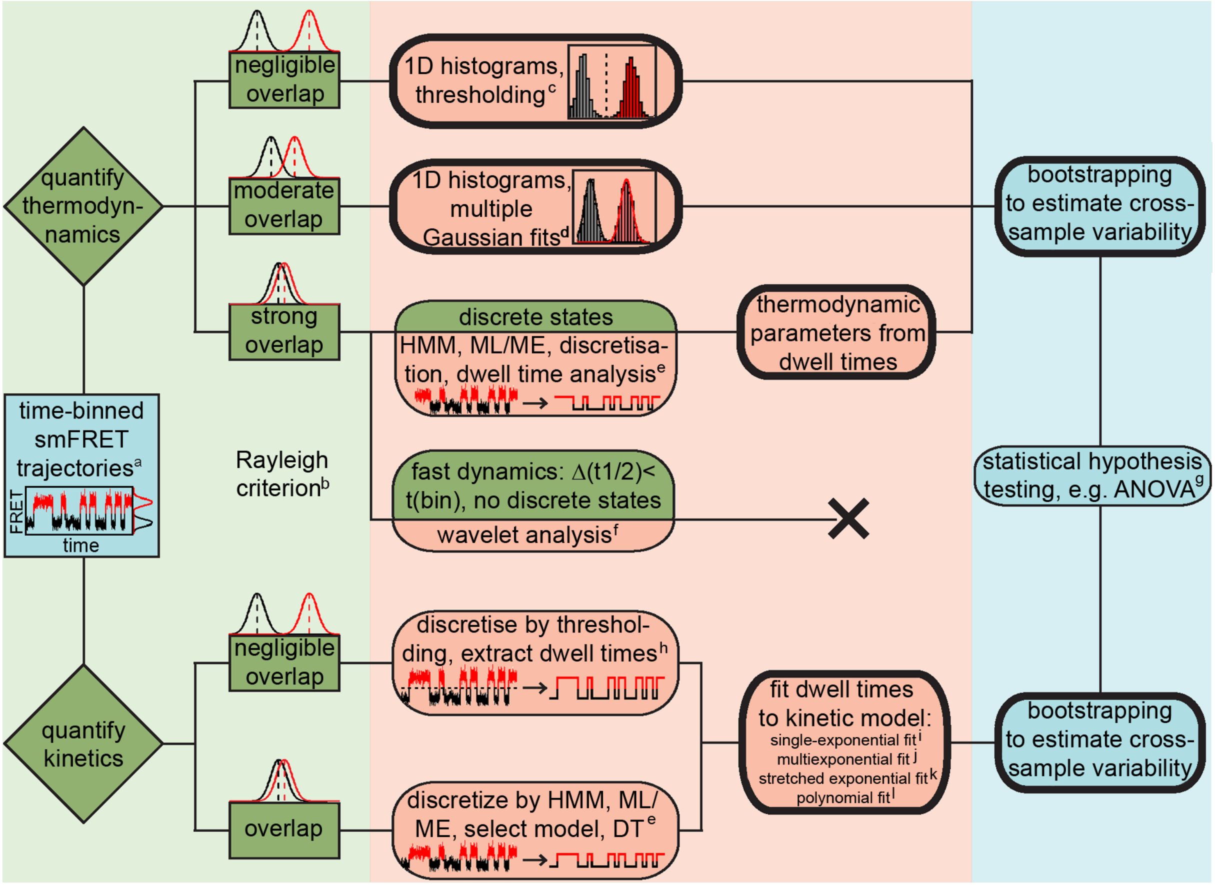


Figure 8.1: Generalised scheme for analysing time-binned smFRET data. Bootstrapping can be used both in thermodynamic and kinetic analysis and is compatible with numerous data formats. Bold frames indicate functionalities available in BOBA FRET. a) As defined in the introduction, see also Gopich and Szabo [436]. b) Rayleigh criterion: two subpopulations are indistinguishable when their peak positions are separated by one standard deviation or less [121]. c) See [434]. d) See [435, 204, 87]. e) See [432, 192, 394, 397, 437, 438, 439]. f) See [440, 441]. g) Multivariate tests (MANOVA) are conceivable to assess whether two or more outcome variables are significantly different at a time, for example the center and the width of a FRET distribution [429]. h) See [197]. i) j) See [204]. k) See [195, 428]. l) Typically used in fluorescence correlation spectroscopy (FCS) [442].

recently been implemented to denoise smFRET data [445]. An extensive overview of strategies for noise removal in so-called piecewise constant signals (constant signal levels connected by abrupt transitions) has been given elsewhere [444].

Hidden-Markov modeling (HMM, Figure 8.1) was first applied on TCSPC data by Yang and Xie [446, 447], and later utilised for analysing time-binned FRET trajectories by the groups of Ha (*HaMMy*, [192]), Gonzalez Jr. (*vbFRET*, [394]), Herschlag (*SMART*, [397]), and Dillingham (*CSSR*, [437]), as well as groups from other research fields (*QuB*, [438]). Briefly, a Markov process is a sequence of state-to-state transitions, becoming hidden because of the experimental noise [192]. Consequently, HMM attempts to reconstruct the underlying time trace based on transition probabilities of a molecule from a state $A$ to a state $B$, and emission probabilities, *i.e.* the likelihood of observing a $FRET$ value when the system is in a discrete state $l$ assuming the noise can be modeled by a given statistical distribution [435, 448]. Different approaches have been employed to determine the exact number of states: (i) deliberate overfitting followed by model selection using the Bayesian information criterion (BIC) or the Akaike information criterion (AIC) [192, 394, 437], or (ii) a maximum evidence approach for both model selection and determination of the model parameters [394]. Hidden Markov approaches enjoy great popularity nowadays such that an extensive body of literature has been published on this topic, including implementations for short time traces [449, 450] and multivariate HMM dealing with more than one time trace at a time [432, 439]. Nonetheless, it should be mentioned that the basic assumptions do not always hold true for single-molecule processes (single-exponential kinetics, *vide infra*), especially when memory effects or large variations in folding kinetics are observed that go beyond the scope of classical kinetics [164, 440].

With the cumulated histograms and/or the dwell times at hand, both the thermodynamic equilibrium and the kinetics associated with the conformational changes can be characterised. To this end, the corresponding error is typically estimated via the goodness of the fit to the data (GOF) [87, 204]. The GOF reports on how well the model describes the experimental data and is mainly determined by the SNR. Important contributions to the noise are made by the stochastic nature of photon emission (shot-noise), background noise, electron multiplier noise, read-out noise, dark noise, resolution-induced noise [189, 436, 451, 452, 453, 454, 455], as well as photophysical effects like quantum yield fluctuations and spectral changes or technical aberrations such as focal drift or fluctuations in laser intensity [146, 189, 455]. In turn, this approach neglects cross-sample variability (differences between single molecules) as it relies on building an ensemble from all smFRET time traces at once. Single-molecule data are however known to frequently display intermolecular heterogeneities that may originate from limitations with regard to the observation time (photobleaching) or technical issues. These frequently manifest as pronounced differences regarding the relative population of conformational states, and as differences in the absolute $FRET$ values observed between individual time traces (*heterogeneous broadening*) [163, 164]. Consequently, approximation of the error by the GOF is expected to underestimate the variance at the expense of the robustness of data interpretation.

It must be emphasised that precise estimation of the variance of the sample is crucial in order to assess whether a difference between different treatment groups is real or has occurred solely by chance, for example by a change in the relative population of the conformational states in response to the addition of a small molecule. Such statistical testing has, to the best of our knowledge, not been reported in the field of single-molecule FRET.

Pioneered by Efron [456], the bootstrap scheme is a resampling method to assess the accuracy of sample estimates that has since been applied in numerous branches of biological research including phylogenetics [457], environmental science [458], force-based single-molecule biophysics [337, 459], or molecular dynamics simulations in conjunction with smFRET experiments on freely diffusing molecules [460]. In bootstrapping, the distribution of the whole population, including measures of variance, is estimated from a sample distribution of the size $n$ ($n$ = number of replicates) [461]. During the resampling process, $N$ values of the sample distribution are randomly selected with an equal probability of $1/n$ and multiple selections are allowed (resampling with replacement) [461]. Typically, $N = n$ to avoid pseudoreplication and the resampling procedure is repeated $M$ times to compute the variance, where $100 \leq M \leq 500$ is usually considered sufficiently robust in phylogenetic research, though more conservative approaches may involve several thousand rounds of bootstrapping [457].

To meet the challenge of making smFRET data analysis more robust, we have designed a software package called BOBA FRET (BOotstrap-BAsed analysis of smFRET data) to estimate the cross-sample variability associated with time-binned smFRET measurements using Efron's bootstrap (Figure 8.1) [456]. The program is freely available, its implementation is straightforward. Herein, we illustrate its workflow to perform both thermodynamic and kinetic analysis of smFRET data: First, the algorithm is shown to be compatible with well-established approaches to analyse smFRET time traces and we characterise its robustness using a set of simulated data. Second, BOBA FRET is applied to an experimental dataset, the cation-dependent interaction of the exon-binding sequence 1 (d3'EBS1*) and the intron-binding sequence 1 (IBS1*), which are derived from a crucial part of the 5'splice site recognition complex in the group II intron *Sc*.ai5$\gamma$ found in *Saccharomyces cerevisiae* (Figure 2). With the bootstrapped errors at hand, we perform statistical hypothesis testing to assess whether cation-induced effects on interaction kinetics and the position of the conformational equilibrium are statistically significant [50, 51, 388].

## 8.2 Materials and Methods

### 8.2.1 Simulations

smFRET time traces were simulated for an intramolecular two-state system. First, discretised time traces were created under the assumption that state-to-state transitions are governed by single-exponential kinetics, followed by addition of Gaussian noise. Standard parameters were based on previous simulations and defined as follows: $FRET_{\mathrm{A}} = 0.3$ (undocked state), $FRET_{\mathrm{B}}$

= 0.7 (docked state); *SNR* = 3.5 (average total intensity = 24.5 photons bin$^{-1}$ s$^{-1}$); *SNR distribution width* = 0; *observation time* = 4000 s; $k_{\text{docking}}$ = 0.1 s$^{-1}$, $k_{\text{undocking}}$ = 0.04 s$^{-1}$ (average number of transitions = 114 per time trace) [192]. For each set of parameters, 100 time traces were analysed, followed by an estimation of the cross-sample variability (*vide infra*). All simulations were performed using a home-built script written in Matlab.

### 8.2.2 Oligonucleotides

The RNA sequence pair was derived from the exon-binding site 1 (EBS1) and the intron-binding site 1 (IBS1) found in the primary *cox1* transcript in *S. cerevisiae*. They are referred to as d3'EBS1* and IBS1* according to the nomenclature used in previous studies (Figure 8.2) [50, 388]. Labeled oligonucleotides were purchased PAGE-purified from IBA AG (Göttingen, Germany) and additionally HPLC purified [426]. All chemicals were purchased from Sigma-Aldrichs (Buchs, Switzerland).

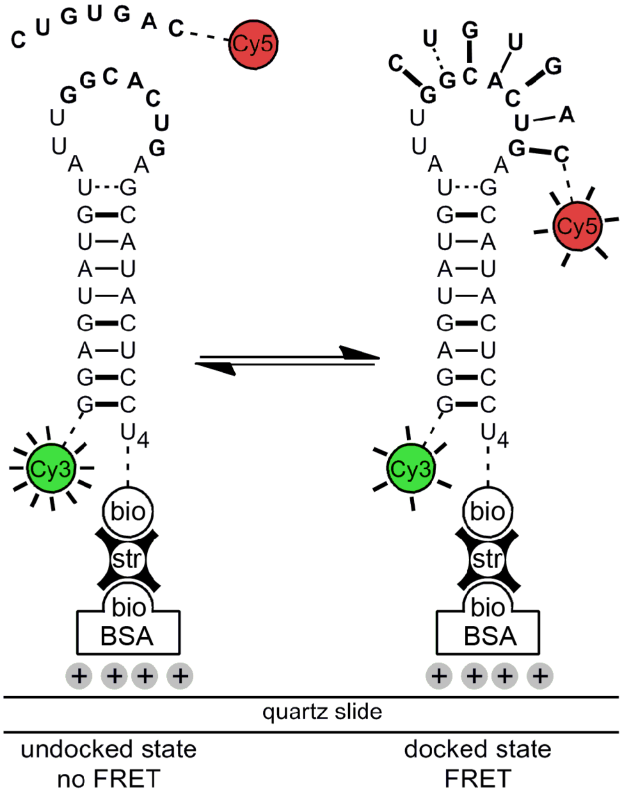


Figure 8.2: Studying d3'EBS1*/IBS1* interaction by smFRET. The d3'EBS1* hairpin is labeled with Cy3 and tethered to the surface of a quartz slide passivated with biotinylated BSA via a biotin-streptavidin linkage. Docking of a Cy5-IBS1* strand is characterised by the appearance of Cy5 fluorescence and a decrease in Cy3 emission due to FRET. Figure adapted from [413].

### 8.2.3 smFRET imaging

Microfluidic channels for total internal reflection microscopy (TIRFM) were prepared from quartz slides (Finkenbeiner, Waltham, MA, USA) as described [290]. The inner surface of the chamber was passivated with biotinylated BSA (Sigma-Aldrich, Buchs, Switzerland), and Cy3-labeled d3'EBS1* was immobilised via a biotin-streptavidin linkage (Figure 2) [170]. The smFRET imaging buffer contained 50 mM MOPS, 100 mM $KNO_3$, 1 mM $M(NO_3)_2$ ($M^{2+}$ = $Ni^{2+}$ or $Co^{2+}$), 1 % D-glucose, 165 U/mL glucose oxidase, 2170 U/mL catalase, 1 mM Trolox, 25 nM Cy5-labeled IBS1*, pH 6.90 [412]. Cy3 and Cy5 emission levels were monitored in a prism-based total internal reflection fluorescence microscope upon alternating laser excitation (ALEX) as described elsewhere [170, 354]. Briefly, fluorophores were excited at 532 and 640 nm in an alternating fashion using diode lasers (CrystaLaser lc., Reno, NV, USA) attenuated to an intensity of $\sim$ 5 mW using neutral density filters (Laser2000 GmbH, Wessling, Germany). Fluorophore emission was spectrally separated with dichroic mirrors (AHF AG, Tübingen, Germany) and projected side-by-side onto a CCD camera (Andor Technology plc., Belfast, Northern Ireland). Photons were collected over 6 minutes at a spatial resolution of 256x256 pixels and a time resolution of 100 ms.

### 8.2.4 Data analysis

smFRET movies were analysed with a home-built Matlab software (Matlab version 8.20.701, license 49040, MathWorks, Nattick, MA, USA). Briefly, the local level of background noise was determined and subtracted from dye emission profiles by creating a sub-image (20x20 pixel), followed by calculating the mean photon count rate of the 20 darkest pixels within this area, a method to locally determine background noise adapted from the commonly used aperture photometry approach [189, 224]. Fluorescence time traces were further corrected for leakage of Cy3 emission into the Cy5 channel ($\sim$ 7 %, determined experimentally). Emission time traces were manually selected for anticorrelation and stable acceptor emission to calculate time-dependent apparent FRET efficiencies, $FRET(t)$, as

$$FRET(t) = \frac{PC(t)^{\text{Cy3 exc}}_{\text{Cy5 em}}}{PC(t)^{\text{Cy3 exc}}_{\text{Cy3 em}} + PC(t)^{\text{Cy3 exc}}_{\text{Cy5 em}}} \tag{8.1}$$

where $PC(t)^{Cy3\ exc}_{Cy3\ em}$ denotes the Cy3 photon count rate upon Cy3 excitation, and $PC(t)^{Cy3\ exc}_{Cy5\ em}$ stands for Cy5 emission upon Cy3 excitation.

### 8.2.5 Characterisation of the thermodynamic equilibrium

To characterise the thermodynamic equilibrium, $n$ individual FRET time traces $FRET(t)_i$ were binned to 1D histograms $\tau_i(FRET)$ using a binning interval of 0.01 FRET units, yielding $m$ individual FRET bins. Subsequently, a normalised cumulated $FRET$ histogram was created for all smFRET data recorded under identical imaging conditions:

$$\bar{\theta}(FRET) = \frac{\sum_{i=1}^{n} \theta_i(FRET)}{\sum_{i,j=1}^{n,m} \theta_i(FRET_j)}, i = 1,2,...,n \text{and} j = 1,2,...,m. \tag{8.2}$$

While individual time traces may be inconclusive in some cases depending on the observation time, the conformational interconversion kinetics, the SNR and the complexity of the system, distinct *FRET* distributions will develop in the cumulated *FRET* histogram if discrete conformational species are present and resolvable [189]. The relative occurrence of these states was then quantified by thresholding or multiple Gaussian fitting (Equations 8.17 and 8.18). In threshold-based analysis, the occurrence is quantified by the integral over the area of the cumulated *FRET* histogram that is assigned to one conformation. For this purpose, the integration limits are defined as -∞, $th_1$, ..., $th_n$, +∞, where *th* refers to a threshold. Without a loss of generality, we defined the threshold value to distinguish two FRET distributions A and B as ($FRET_A$ + $FRET_B$)/2, which corresponds to the midpoint between their centers $FRET_A$ and $FRET_B$.
Characterisation of the thermodynamic equilibrium was also performed using dwell times. The underlying principle is that the time the molecules spend in different discrete states can be directly used to infer the position of the conformational equilibrium. For d3'EBS1*/IBS1*, the docked fraction was used to calculate the association constants $K_\mathrm{A}$ from the FRET histograms:

$$K_\mathrm{A} = \frac{c(\text{d3'EBS1*-IBS1*})}{c(\text{d3'EBS1*})c(\text{IBS1*})} \simeq \frac{p(FRET_\text{high})}{p(FRET_\text{zero})c_\text{tot}(\text{IBS1*})} \tag{8.3}$$

assuming that $c$(IBS1*) $>>$ $c$(d3'EBS1*). Here, $c$(d3'EBS1*), $c$(IBS1*) and $c$(d3'EBS1*-IBS1*) denote the concentrations of free and docked oligonucleotides, and $c_\text{tot}$(IBS1*) corresponds to the total concentration of IBS1*. $p(FRET_\text{high})$ and $p(FRET_\text{zero})$ refer to the docked and the undocked fractions as determined from the cumulative probability plots. Association constants were determined from the ensemble of dwell times as

$$K_\mathrm{A} = \frac{1 - \frac{t_\text{zero}}{t_\text{zero}+t_\text{high}}}{\frac{t_\text{zero}}{t_\text{zero}+t_\text{high}} c(\text{IBS1*})_\text{tot}} \simeq \frac{t_\text{high}}{t_\text{zero} c(\text{IBS1*})_\text{tot}} \tag{8.4}$$

with $t_\text{zero}$ and $t_\text{high}$ denoting the total time spent in the undocked and the docked state, respectively. The approaches used to determine dwell times and subsequent processing steps are outlined in the next section.

### 8.2.6 Characterisation of kinetics

Dwell times were determined from individual time traces $FRET(t)_i$ via thresholding at ($FRET_\mathrm{A}$ + $FRET_\mathrm{B}$)/2 or using the freely available software vbFRET [394]. In short, vbFRET employs a maximum evidence (ME) approach for model selection (the number of FRET states *L*), followed by inferring the model parameters (*FRET* values and transitions) by a combination of variational Bayesian expectation maximization and hidden Markov modeling (HMM) [462]. As their duration was unknown, the first and the last dwell time of each time trace were consistently discarded. Additionally, a weighted *k*-means algorithm was applied to transition den-

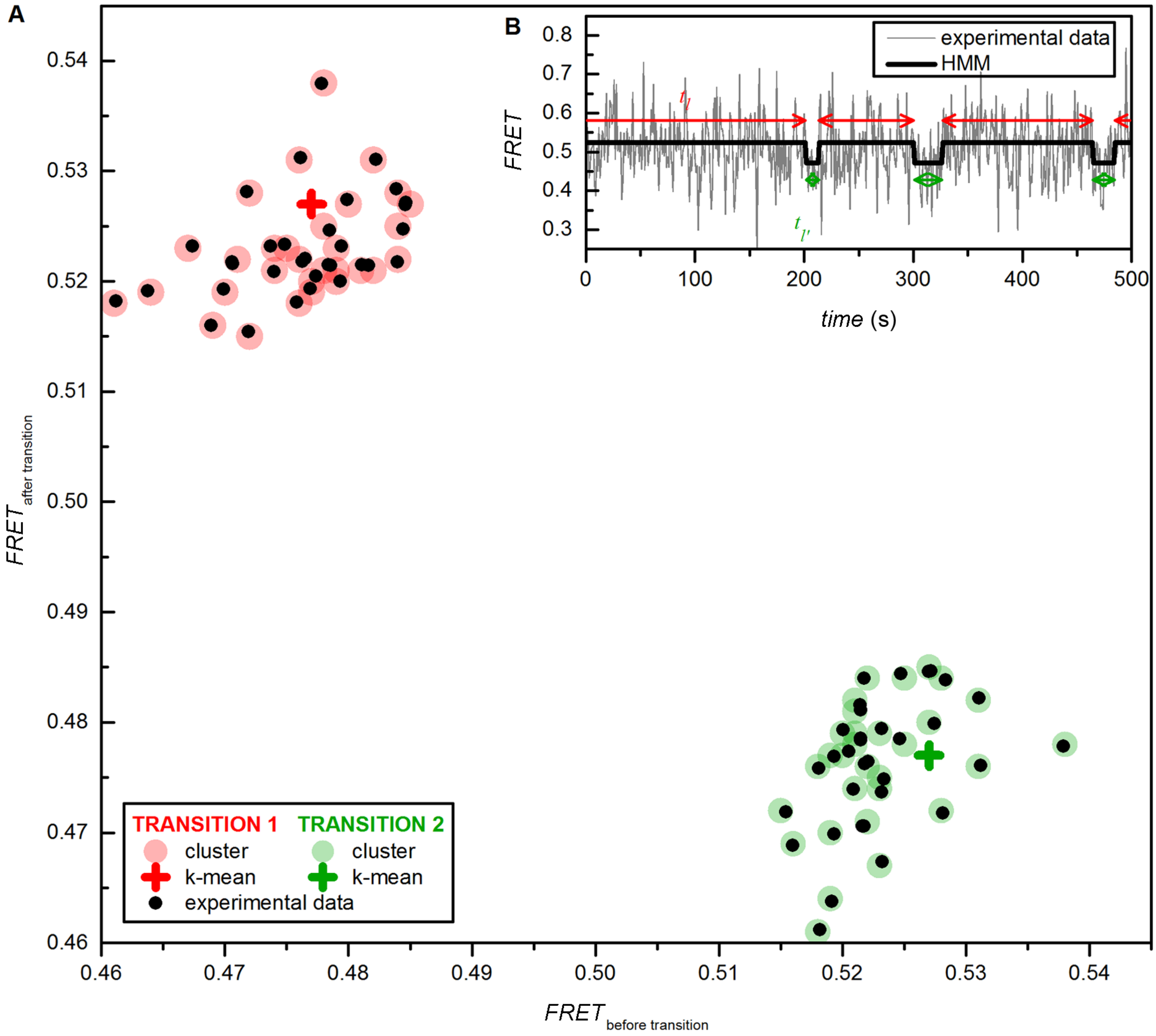


Figure 8.3: *k*-means clustering to assign dwell times to consistent FRET values for further processing steps. (A) Transition density plot (TDP) built from a set of HMM-discretised FRET time traces. The data points are iteratively assigned to one of the two centers according to their distance. The center coordinates are then recalculated according to the distances and occurrences (weights) of the clustered data point. The weighted *k*-mean centers are assumed to be definitive when the set of clustered transition does not change after an additional round of iteration. (B) Dwell time analysis of one simulated FRET time trace for a two state system: $\Delta\text{FRET} = 0.04$, $FRET_A = 0.48$ (undocked state), $FRET_B = 0.52$ (docked state); $SNR = 6.0$ (width $\sigma = 0.143$); *observation time* = 4000 s (magnified to highlight transitions); $k_{\text{docking}} = 0.04\ \text{s}^{-1}$ (intramolecular reaction) $k_{\text{undocking}} = 0.1\ \text{s}^{-1}$. Each of the two *FRET* states detected in the trace are assigned to the center of one of the two clusters and the corresponding dwell times are subsequently used for thermodynamic or kinetic analysis.

sity plots (TDP) created from the vbFRET data to cluster the coordinates ($FRET_{\text{before transition}}$; $FRET_{\text{after transition}}$) into $k$ subgroups and assigned each transition to one of the $k$ centers ($\langle FRET_{\text{before transition}} \rangle_k$, $\langle FRET_{\text{after transition}} \rangle_k$). The principle of $k$ means clustering is illustrated in Figure 8.3 and is a well-precedented approach to cluster data that has been applied to heterogeneous HMM data previously [463, 464].

For single-exponential state-to-state transitions occurring in a stochastic manner with rate constants that do not vary over time, $k$ subgroups in the TDP correspond to $L$ $FRET$ states with $k = L^2 - L$. The corresponding dwell times are in this case exponentially distributed [397]. Consequently, dwell times were binned to histograms that then were used to calculate the normalised cumulative probability distributions *normalised cumP* , which were in turn fitted to exponential decay functions to extract the corresponding rate constants [196, 197, 219]. Here, single- and stretched exponential decays were used to approximate simulated and experimental data [195, 198, 204]:

$$1 - normalised\ cumP = \sum_{p=1}^{O} a_p \exp\left\{-(\frac{t}{\tau_p})\right\}, p = 1, 2, ..., O, \tag{8.5}$$

$$1 - normalised\ cumP = \exp\left\{-(\frac{t}{\tau_{1/\mathrm{e}}})^{\beta}\right\} \tag{8.6}$$

where $O$ denotes the number of exponential decays (single-exponential: $O = 1$), $a_p$ is the amplitude, and $\tau_p$ the average dwell time in the conformational state (decay constant). The decay time $\tau_{1/e}$ refers to the time required for *1 - normalised cumP* to drop to 1/e of its initial value and the stretching exponent $\beta$ ($0 < \beta \leq 1$) is a means to quantify the width of the rates distribution [428]. Both $\tau_p$ and $\tau_{1/e}$ were used to determine the rate constants associated with conformational changes:

$$k_{\text{docking}} = \frac{1}{\tau_{\text{docking}} c_{\text{tot}}(\text{IBS1*})} \tag{8.7}$$

$$k_{\text{docking}} = \left(\left(\frac{\tau_{1/\text{e,undocking}}}{\beta}\right)\Gamma\left(\frac{1}{\beta}\right)\right)^{-1} \tag{8.8}$$

for a first order reaction, where $\Gamma$ is the so-called Gamma function [428].

### 8.2.7 Bootstrapping in thermodynamic and kinetic analysis of smFRET data

Following the conventions in the field, the variability of the data vector is assumed to be due to limited observation time, experimental noise, instrumental aberrations (*heterogeneous broadening*, *vide supra*), and irresolvable molecular motion [192, 394]. Bootstrapping allows to characterise the data space of an ensemble of smFRET time traces, and thus, to quantify cross-sample variability and allowing for its application in statistical hypothesis testing.

Bootstrap samples were built for a multi-sample problem given by a random sample of $n$ smFRET time traces, each of which is composed of a discrete number of time bins $B$ $\{FRET(t)_1$, $FRET(t)_2$, ..., $FRET(t)_n\}$, observed from a completely unspecified probability distribution $F$ according to Efron [456]. The ensemble of time trajectories were used to create the corresponding single molecule *FRET* histograms $\{\theta_1(FRET)$, $\theta_2(FRET)$, ..., $\theta_n(FRET)\}$. Resampling was then performed with replacement, where each single-molecule time trace $FRET(t)_i$ has a probability of

$$\frac{B_n}{\sum\limits_n B_n} \tag{8.9}$$

to be selected. Here, $B_n$ denotes number of time bins of the $n$th individual *FRET* time trace, and the whole expression can be regarded as a weighting factor that accounts for differences in length of individual time traces. Subsequently, bootstrap samples (boba) were created from previously selected time traces $FRET(t)_i^*$ and *FRET* histograms $\theta_i^*(FRET)$:

$$FRET(t)_{\text{boba}} = (FRET(t)_1^*, FRET(t)_2^*, ..., FRET(t)_N^*), i = 1, 2, ..., N \tag{8.10}$$

$$\theta(t)_{\text{boba}} = (\theta_1^*(FRET), \theta_2^*(FRET), ..., \theta_N^*(FRET)), i = 1, 2, ..., N \tag{8.11}$$

where $N$ was set to $n$ to prevent pseudoreplication [457]. It should be emphasised, that in the case of an equal length of the time traces (constant observation time, no photobleaching etc.) the probability simplifies to $1/n$, *i.e.* each time trace and its corresponding histogram has the same probability of being selected (molecular weighting). Normalised cumulated FRET histograms of the bootstrap-based ensemble were calculated as:

$$\bar{\theta}(FRET)_{\text{boba}} = \frac{\sum\limits_{i=1}^{N} \theta_i^*(FRET)}{\sum\limits_{i,j=1}^{N,m} \theta_i^*(FRET_j)} \tag{8.12}$$

using a Monte Carlo method to approximate the bootstrap distribution with a random sample of the size $N$, the creation of bootstrap samples was repeated $M$ times, yielding an independent random ensemble of bootstrap time traces $FRET(t)^1_{\text{boba}}$, $FRET(t)^2_{\text{boba}}$, ..., $FRET(t)^M_{\text{boba}}$, as well as the corresponding histograms $\theta(FRET)^1_{\text{boba}}$, $\theta(FRET)^2_{\text{boba}}$, ..., $\theta(FRET)^M_{\text{boba}}$ and normalised cumulated *FRET* histograms $\bar{\theta}^1_{\text{boba}}$, $\bar{\theta}^2_{\text{boba}}$, ...$\bar{\theta}^M_{\text{boba}}$. The bootstrap mean $\bar{X}_{\text{boba}}$ and the corresponding standard deviation $\sigma_{\text{boba}}$ were estimated according to [461]:

$$\sigma_{\text{boba}} = \sqrt{\sum_{k=1}^{M} \frac{(X_k - \bar{X})^2}{M-1}}, k = 1, 2, ..., M. \tag{8.13}$$

Here, $X$ denotes the random parameter whose variability is to be estimated, for example the relative occurrence of a certain *FRET* population $A_l$ given by a certain state $l$ in the thermodynamic analysis.

The bootstrap distribution of $X_{\mathrm{boba}} = X(FRET(t)_{\mathrm{boba}}, \hat{F})$, depends on both the random sample $FRET(t)_{\mathrm{boba}}$ and the sample probability distribution $\hat{F}$. $X_{\mathrm{boba}}$ is expected to approximate the real underlying distribution $X(FRET(t), F)$ well, including its mean and standard deviation. In this study, we chose $M = 100$, following the conventions from other fields [457], because a time-consuming increase of $M$ would yield only moderate improvements (Figure 8.4) [456]. It is important to emphasise that the noise-induced fluctuation around discrete values in smFRET

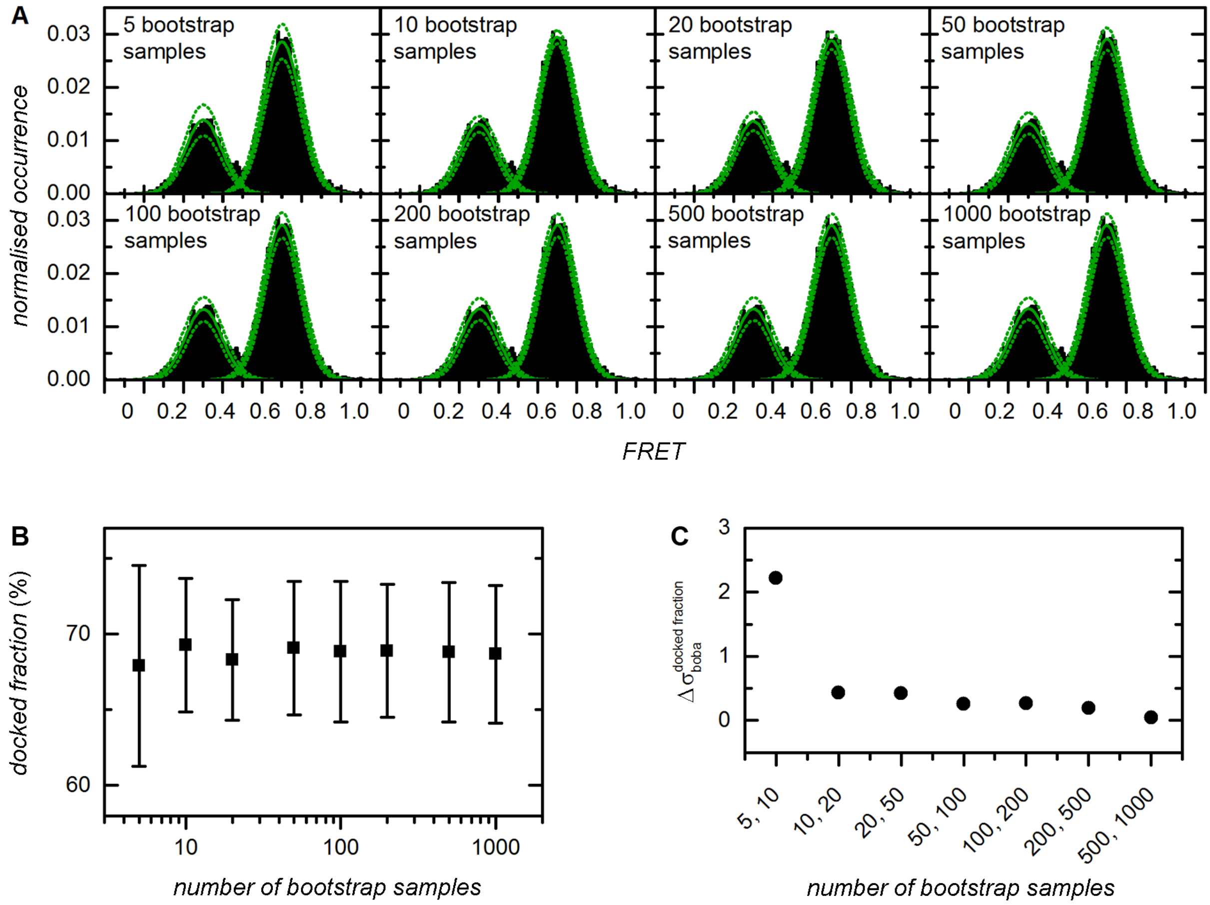


Figure 8.4: Dependence of the bootstrapped estimated cross-sample variability on the number of bootstrap samples. (A) Gaussian fitting was performed in conjunction with bootstrapping to analyse 100 simulated smFRET time traces ($N = 100$, Equation 8.11). The number of bootstrap samples was varied between 5 and 1000 ($M$, Equation 8.13). The histogram corresponds to the normalised cumulated histogram built from all time traces (Equation 8.2), solid lines depict Gaussian fit functions, dashed lines the variability associated with the amplitude and the width ($3\sigma_{\mathrm{boba}}$). (B) Fraction of docked molecules and cross-sample variability, data from panel (A). Error bars correspond to $3\sigma_{\mathrm{boba}}$. (C) Dependence of $\Delta\sigma_{\mathrm{boba}}$ on the number of bootstrap samples. Data point correspond to the difference in $3\sigma_{\mathrm{boba}}$ of adjacent data points and demonstrate that fluctuations become negligible when more than 100 bootstrap samples are used. Parameters of the simulation: $FRET_{\mathrm{A}} = 0.3$ (undocked state), $FRET_{\mathrm{B}} = 0.7$ (docked state); $SNR = 3.5$; observation time = 100 s; $k_{\mathrm{docking}} = 0.1\ \mathrm{s}^{-1}$, $k_{\mathrm{undocking}} = 0.04\ \mathrm{s}^{-1}$.

time traces is entirely time-independent (stochastic). This is not always the case for time series, which would then require more sophisticated mathematical treatments (Figure 8.5) [456, 465].

### 8.2.8 Bootstrapping and regression (method 1)

To estimate the bootstrap mean $\bar{X}_{\mathrm{boba}}$ and the standard deviation $\sigma_{\mathrm{boba}}$ of the parameter $X$, we defined a reasonably general non-linear regression model:

$$y_j = g_j(\alpha, x_j) + \varepsilon_j, j = 1, 2, ..., m \tag{8.14}$$

where $g$ denotes a model function of the unknown parameter vector $\alpha$ approximating the data vector $y$ (outcome variable) depending on $x$ (input variable), both of which display the length $m$. The corresponding residuals $\varepsilon_j$ follow the unspecific probability distribution $\varepsilon_j \sim F$. We

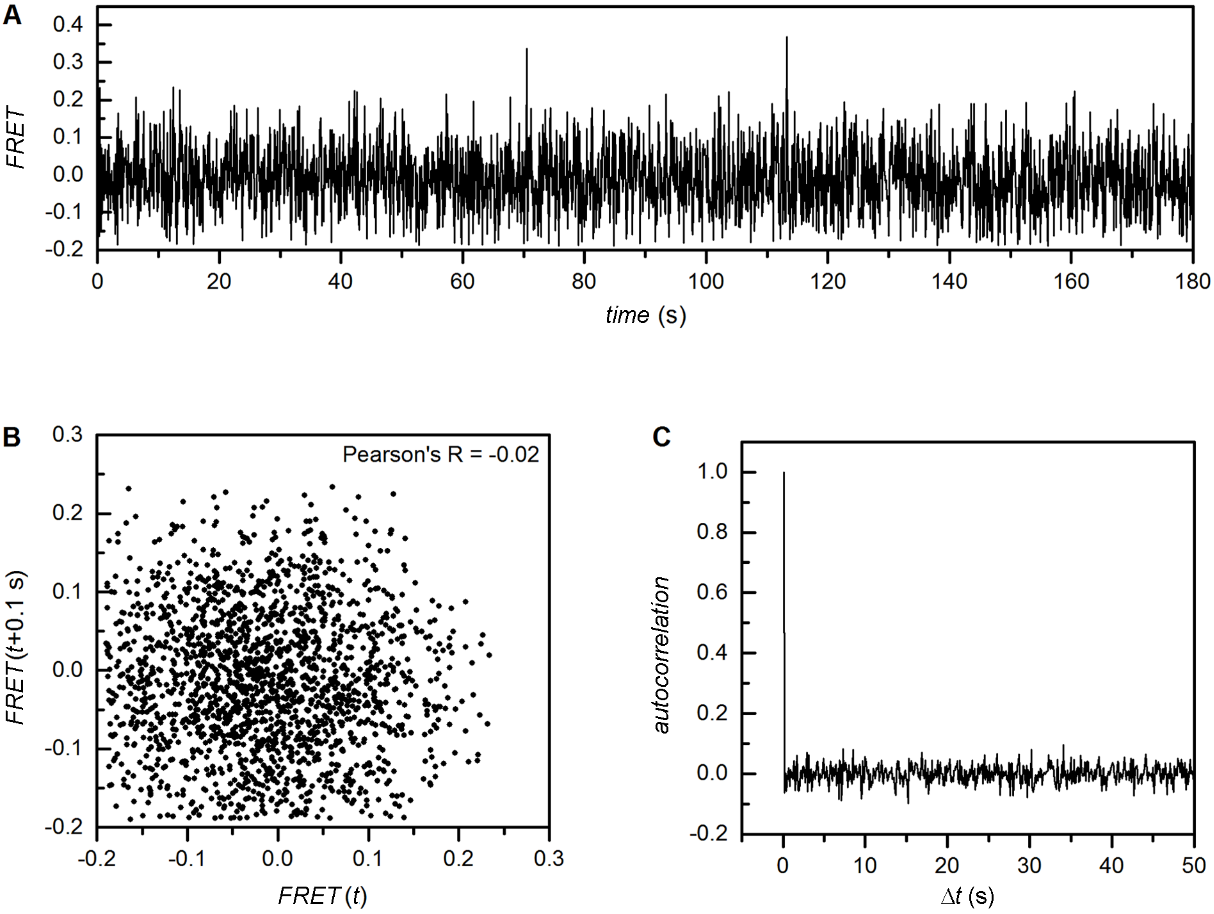


Figure 8.5: Statistical nature of noise in smFRET data. (A) Cy3 emission time trace, representative data. Surface-tethered Cy3-tagged d3'EBS1* fluctuates around zero *FRET* in the absence of IBS1*. (B) *FRET*($t$) versus *FRET*($t$ + 100 ms) scatter plot of the data shown in (A) develops as a two-dimensional Gaussian distribution. Time-dependent noise would be expected to accumulate on a diagonal. (C) The autocorrelation function of the data in (A) clearly demonstrates that the noise of the time trace shown in (A) is time-independent.

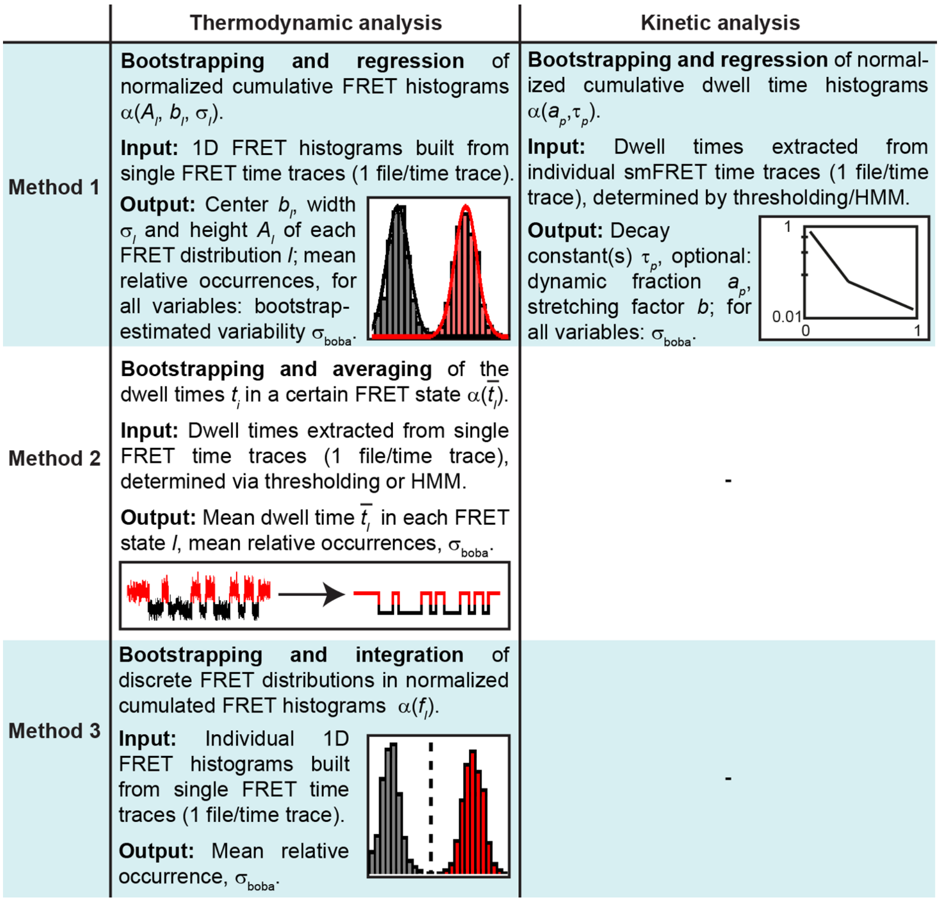
fitted $y$ based on a non-linear least square regression to estimate $\alpha$ [429]:

$$\hat{\alpha} : \min_{\alpha} \sum_{j=1}^{m} \left(y_j - g_j(\alpha, x_j)\right)^2, j = 1, 2, ..., m, \tag{8.15}$$

which yields the sampling distribution of $\hat{\alpha}$. Subsequently, bootstrap samples were generated according to Equations 8.10, 8.11 and 8.12 and are henceforth referred to as $y_{\text{boba}}$ using the terminology of Equation 8.14:

$$y_{\text{boba},j} = g_j(\alpha, x_j) + \varepsilon_j, j = 1, 2, ..., m. \tag{8.16}$$

Regression based on a non-linear least square criterion was performed in an analogous manner as in Equation 8.15:



| | Thermodynamic analysis | Kinetic analysis |
|---|---|---|
| **Method 1** | **Bootstrapping and regression** of normalized cumulative FRET histograms $\alpha(A_l, b_l, \sigma_l)$.<br>**Input:** 1D FRET histograms built from single FRET time traces (1 file/time trace).<br>**Output:** Center $b_l$, width $\sigma_l$ and height $A_l$ of each FRET distribution $l$; mean relative occurrences, for all variables: bootstrap-estimated variability $\sigma_{\text{boba}}$. | **Bootstrapping and regression** of normalized cumulative dwell time histograms $\alpha(a_p, \tau_p)$.<br>**Input:** Dwell times extracted from individual smFRET time traces (1 file/time trace), determined by thresholding/HMM.<br>**Output:** Decay constant(s) $\tau_p$, optional: dynamic fraction $a_p$, stretching factor $b$; for all variables: $\sigma_{\text{boba}}$. |
| **Method 2** | **Bootstrapping and averaging** of the dwell times $t_l$ in a certain FRET state $\alpha(\bar{t}_l)$.<br>**Input:** Dwell times extracted from single FRET time traces (1 file/time trace), determined via thresholding or HMM.<br>**Output:** Mean dwell time $\bar{t}_l$ in each FRET state $l$, mean relative occurrences, $\sigma_{\text{boba}}$. | - |
| **Method 3** | **Bootstrapping and integration** of discrete FRET distributions in normalized cumulated FRET histograms $\alpha(f_l)$.<br>**Input:** Individual 1D FRET histograms built from single FRET time traces (1 file/time trace).<br>**Output:** Mean relative occurrence, $\sigma_{\text{boba}}$. | - |

Figure 8.6: Summary of the different analytical approaches performed in conjunction with bootstrapping to extract thermodynamic or kinetic parameters from time-binned smFRET data in this study. The respective input and output variables are indicated as well. Please refer to the methods section for a detailled mathematical description.

$$\hat{\alpha}_{boba} : \min_{\alpha} \sum_{j=1}^{m} \left(y_{\text{boba},j} - g_j(\alpha, x_j)\right)^2, j = 1, 2, ..., m. \tag{8.17}$$

Applying this procedure on $M$ independent bootstrap samples yielded a random sample $\hat{\alpha}^1_{\text{boba}}$, $\hat{\alpha}^2_{\text{boba}}$, ..., $\hat{\alpha}^M_{\text{boba}}$ that was used to estimate $\hat{X}_{\text{boba}}$ and $\sigma_{\text{boba}}$. These values were later used for analysis of variance (ANOVA) [429].

The non-linear regression model was then applied to the normalised cumulated 1D *FRET* histograms $\bar{\theta}(FRET)$ to quantify the variability associated with the determination of thermodynamic parameters. According to the conventions of the field, different conformational states were quantified by multiple Gaussian fitting:

$$g(\alpha, FRET) = \sum_{l=1}^{L}, A_l \exp\left\{-\frac{(FRET - b_l)^2}{2\sigma_l^2}\right\} = 1, 2, ..., L. \tag{8.18}$$

where $L$ denotes the number of states that was in our case determined beforehand using a maximum evidence approach (*vide supra*), even though other model selection approaches are conceivable [192, 394]. $A_l$ refers to the respective amplitudes, $b_l$ to the center values, and $\sigma_l$ to the width of the distribution. The ensemble of model parameters constitute the parameter vector $\alpha(A_l, b_l, \sigma_l)$. The resulting regression model Equation 8.16 for each bootstrap sample is defined as

$$\overline{\theta}(FRET)_{\text{boba},j} = g_j(\alpha, FRET_j) + \varepsilon_j, j = 1, 2, ...m \tag{8.19}$$

and according to the non-linear least square fitting procedure described in Equation 8.17

$$\hat{\alpha}_{boba} : \min_{\alpha} \sum_{j=1}^{m} \left\{\bar{\theta}(FRET)_{\text{boba},j} - g_j(\alpha, FRET_j)\right\}^2, j = 1, 2, ...m \tag{8.20}$$

we obtained the representation $\hat{\alpha}^k_{boba}(A_l,\ b_l,\ \sigma_l)$ of the sampling distribution $\hat{\alpha}_{boba}$.

Second, we applied the bootstrap-based regression on 1 - *normalised cumP* distributions to quantify the variability associated with the analysis of kinetics (Section 8.2.6, *vide supra*). The appropriate model function based on Equation 8.5 was obtained through the maximum evidence algorithm, which samples the model space as well as the parameter space to find the most evident model and yields the number of components $O$ [394]:

$$g(\alpha, t) = \sum_{p=1}^{O} \exp\left\{-\left(\frac{t}{\tau_p}\right)\right\}, p = 1, 2, ..., O. \tag{8.21}$$

Thus, the regression model Equation 8.16 for each bootstrap sample was defined as:

$$1 - \textit{normalised cumP}(t)_{\text{boba},j} = g_j(\alpha, t_j) + \varepsilon_j, p = 1, 2, ..., m \tag{8.22}$$

and

$$\hat{\alpha}_{boba} : \min_{\alpha} \sum_{j=1}^{m} \left\{1 - normalised\ cumP(t)_{\mathrm{boba},j} - g_j(\alpha, t_j)\right\}^2, j = 1, 2, ..., m \tag{8.23}$$

Thus, we obtained the representation $\hat{\alpha}^k_{\mathrm{boba}}(a_p, \tau_p)$ of the sampling distribution $\hat{\alpha}^k_{boba}(a_p, \tau_p)$ of the sampling distribution $\hat{\alpha}_{boba}$. Considerations regarding method 1 are summarised in Figure 8.6.

### 8.2.9 Bootstrapping and averaging (method 2)

The bootstrapping formalism described above was also applied in the analysis of the thermodynamic equilibrium using dwell times obtained by threshold- or HMM-based analysis of smFRET time traces. Here, each time trace $FRET(t)_i$ is composed of a number of $m$ dwell times $t_{i,j,l}$ in a discrete state $l$. As a consequence, each bootstrap sample $FRET(t)_{\mathrm{boba}}$ yields an average dwell time in a certain state

$$\bar{t}_{\mathrm{boba},l} = \frac{1}{N * m} \sum_{i,j=1}^{N,m} t_{i,j,l} \tag{8.24}$$

where $i$ = 1, 2, ..., $N$ accounts for the time traces of $N$ different molecules and $j$ = 1, 2, ..., $m$ for the dwell times in the state $l$. Again, applying this procedure on $M$ independent bootstrap samples yielded a random sample $\hat{\alpha}^1_{boba}$, sample $\hat{\alpha}^2_{boba}$, ..., sample $\hat{\alpha}^M_{boba}$ that was used to estimate $\bar{X}_{\mathrm{boba}}$ and $\sigma_{\mathrm{boba}}$ of the thermodynamic parameters. Here, we determined the relative occurrence of each state, as well as the equilibrium constant $K_{\mathrm{eq}}$ or, in the special case of an intermolecular association of the type A + B $\leftrightarrow$ AB, the binding constant $K_{\mathrm{A}}$ (Equations 8.3 and 8.4). Considerations regarding method 2 are summarised in Figure 8.6.

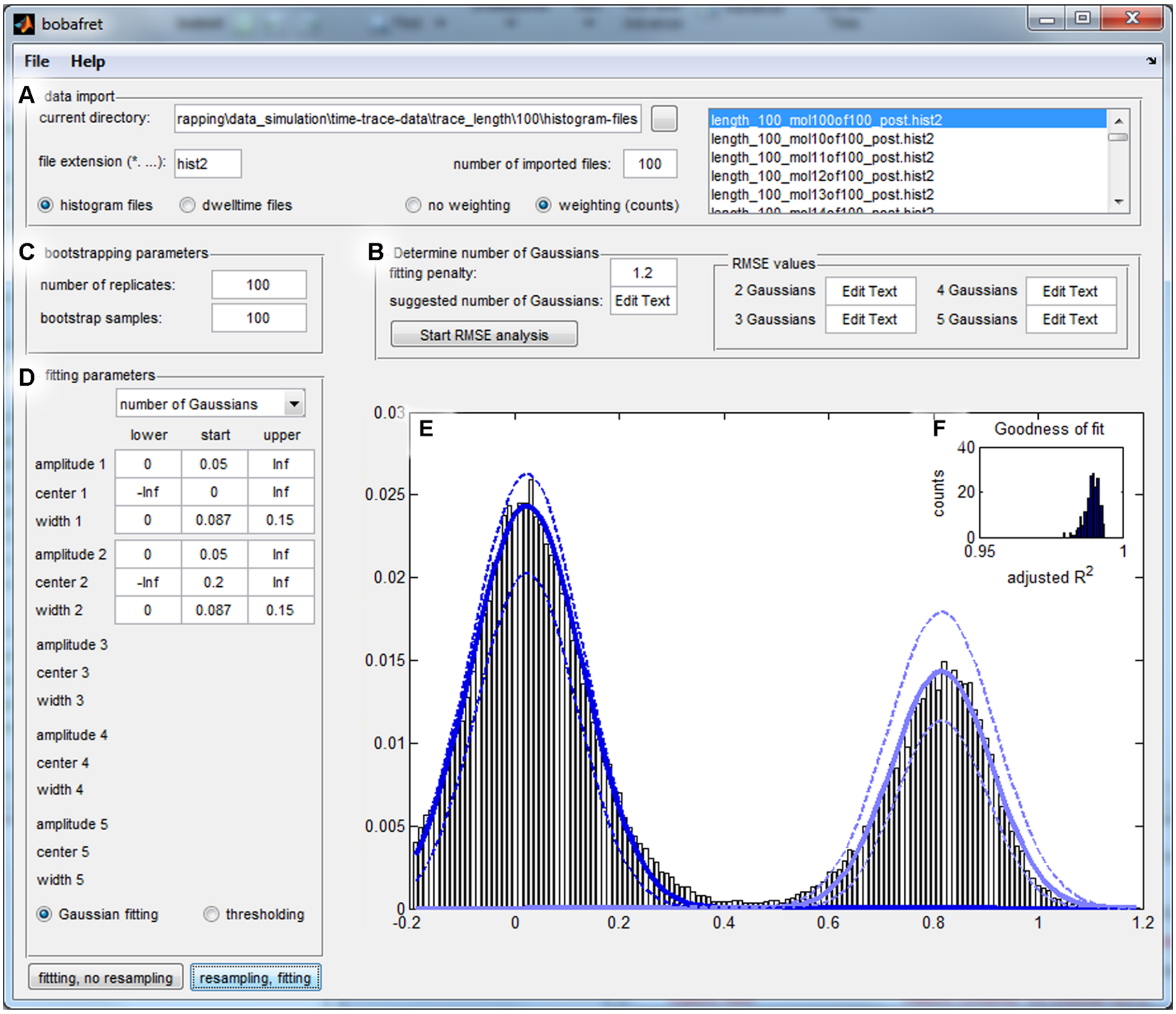
### 8.2.10 Bootstrapping and integration (method 3)

Finally, we applied bootstrapping on normalised cumulated *FRET* histograms $\bar{\theta}(FRET)$ in conjunction with thresholding. Here, each bootstrap sample $\bar{\theta}(FRET)_{\mathrm{boba}}$ yielded a threshold value ($FRET_{\mathrm{boba,A}}$ + $FRET_{boba,B}$)/2 which was used to quantify the relative occurrence of each FRET state as explained before. In an analogous manner, applying this procedure on $M$ independent bootstrap samples allowed us to estimate $\bar{X}_{\mathrm{boba}}$ and $\sigma_{\mathrm{boba}}$ of the relative occurrence of the *FRET* states. These values were later used for analysis of variance (ANOVA) [429]. Considerations regarding method 3 are summarised in Figure 8.6.

Resampling and fitting was done with the software package BOBA FRET that is freely available via http://www.aci.uzh.ch/rna/. The graphical user interface is illustrated in Figure 8.7 and 8.8. Please refer to the Appendix for the complete source code of the software.

## 8.3 Robustness of the software and simulated data

The robustness of the algorithm and its compatibility with common approaches used for thermodynamic and kinetic analysis was assessed using a simple intramolecular two-state system. Normally distributed noise was added to simulated time traces that were varied in length, sep-



current directory: rapping\data_simulation\time-trace-data\trace_length\100\histogram-files

file extension (*. ...): hist2

number of imported files: 100

● histogram files ○ dwelltime files ○ no weighting ● weighting (counts)

length_100_mol100of100_post.hist2
length_100_mol10of100_post.hist2
length_100_mol11of100_post.hist2
length_100_mol12of100_post.hist2
length_100_mol13of100_post.hist2

fitting penalty: 1.2

suggested number of Gaussians: Edit Text

number of replicates: 100

bootstrap samples: 100

number of Gaussians

| | lower | start | upper |
|---|---|---|---|
| amplitude 1 | 0 | 0.05 | Inf |
| center 1 | -Inf | 0 | Inf |
| width 1 | 0 | 0.087 | 0.15 |
| amplitude 2 | 0 | 0.05 | Inf |
| center 2 | -Inf | 0.2 | Inf |
| width 2 | 0 | 0.087 | 0.15 |
| amplitude 3 | | | |
| center 3 | | | |
| width 3 | | | |
| amplitude 4 | | | |
| center 4 | | | |
| width 4 | | | |
| amplitude 5 | | | |
| center 5 | | | |
| width 5 | | | |

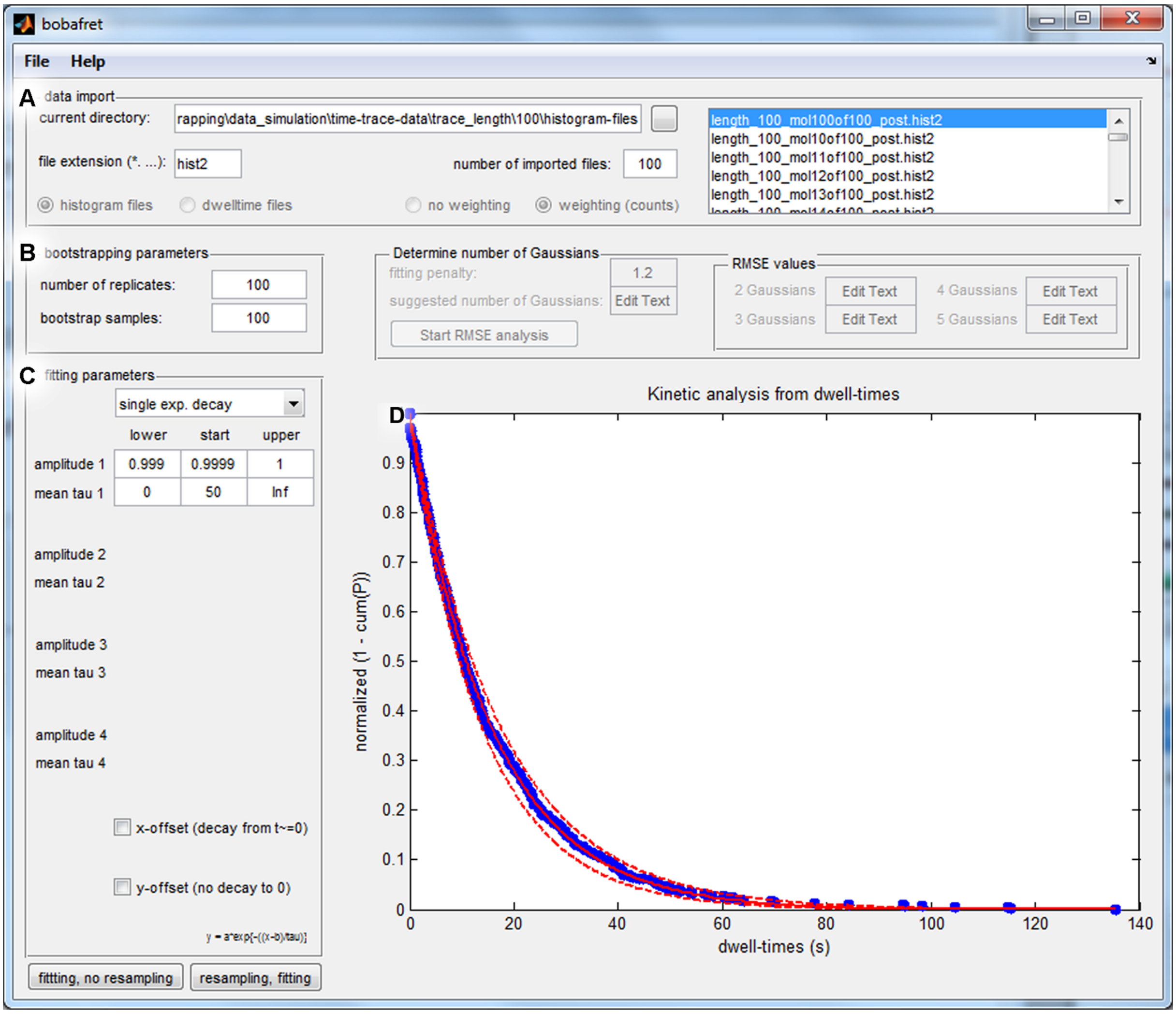
● Gaussian fitting ○ thresholding

Figure 8.7: Boba FRET user interface for thermodynamic analysis. (A) Data import from ASCII files. Both smFRET histogram files (first column: *FRET*, second column: occurrence (counts); further columns are ignored) and dwell time files are supported (first column: duration, second column: *FRET* before transition, third column: *FRET* after transition). (B) Optional determination of the optimal number of Gaussians by distribution analysis [466, 467]. (C) Setting the parameters for bootstrapping ($N$ and $M$, Equations 8.12 and 8.13). (D) Setting the starting guesses and boundaries of the Gaussian fits (Equation 8.18). Alternatively, thresholding can be performed. (E) Original normalised data and fitting results. Solid lines correspond to the fit to the original data, dashed lines to the bootstrapped estimated variability (highest and lowest values of the amplitude and the width). (F) Goodness of fit to all bootstrapped histograms. All fitting parameters (in the case of Gaussian fitting) and the relative occurrences are automatically exported to text files for further analysis.

aration of the *FRET* populations, ratio of the rate constants associated with conformational interconversion, and SNR (Figure 8.9).

## 8.3.1 Thermodynamic characterisation of simulated smFRET data

The relative population of *FRET* states was quantified using four commonly used approaches: Gaussian fitting of normalised cumulated *FRET* histograms (**method 1**), the ratio of dwell times obtained by either thresholding or HMM (both **method 2**) [394], and fractional integra-



data import
current directory: rapping\data_simulation\time-trace-data\trace_length\100\histogram-files
file extension (*. ...): hist2
number of imported files: 100
◉ histogram files ○ dwelltime files ○ no weighting ◉ weighting (counts)

length_100_mol100of100_post.hist2
length_100_mol10of100_post.hist2
length_100_mol11of100_post.hist2
length_100_mol12of100_post.hist2
length_100_mol13of100_post.hist2

bootstrapping parameters
number of replicates: 100
bootstrap samples: 100

Determine number of Gaussians
fitting penalty: 1.2
suggested number of Gaussians: Edit Text

RMSE values
2 Gaussians: Edit Text
3 Gaussians: Edit Text
4 Gaussians: Edit Text
5 Gaussians: Edit Text

fitting parameters
single exp. decay

| | lower | start | upper |
|---|---|---|---|
| amplitude 1 | 0.999 | 0.9999 | 1 |
| mean tau 1 | 0 | 50 | Inf |
| amplitude 2 | | | |
| mean tau 2 | | | |
| amplitude 3 | | | |
| mean tau 3 | | | |
| amplitude 4 | | | |
| mean tau 4 | | | |

☐ x-offset (decay from t~=0)
☐ y-offset (no decay to 0)
y = a*exp[-((x-b)/tau)]

Figure 8.8: Boba FRET user interface for dwell time analysis. (A) Data import from ASCII files. File format: first column, duration; second column, *FRET* before transition; third column, *FRET* after transition. (B) Setting the parameters for bootstrapping ($N$ and $M$, Equations 8.12 and 8.13). (C) Setting the starting values and boundaries of the exponential decay function to be used for fitting. Mono-, bi-, tri-, and tetraexponential decays functions are implemented, as well as stretched exponential decays (Equations 8.5 and 8.6). (D) Original normalised data and fitting results. Solid lines correspond to the fit to the original data, dashed lines to the bootstrapped estimated variability (highest and lowest values for the decay constant). All fitting parameters are automatically exported to text files for further analysis.

tion after thresholding of normalised cumulated *FRET* histograms (**method 3**), respectively.

Figure 8.10A demonstrates that the estimation of the docked fraction becomes more accurate at longer observation times. At the same time, the bootstrap-estimated error scales inversely to the length of time traces. This is expected, as longer time traces yield more data points. Dwell-

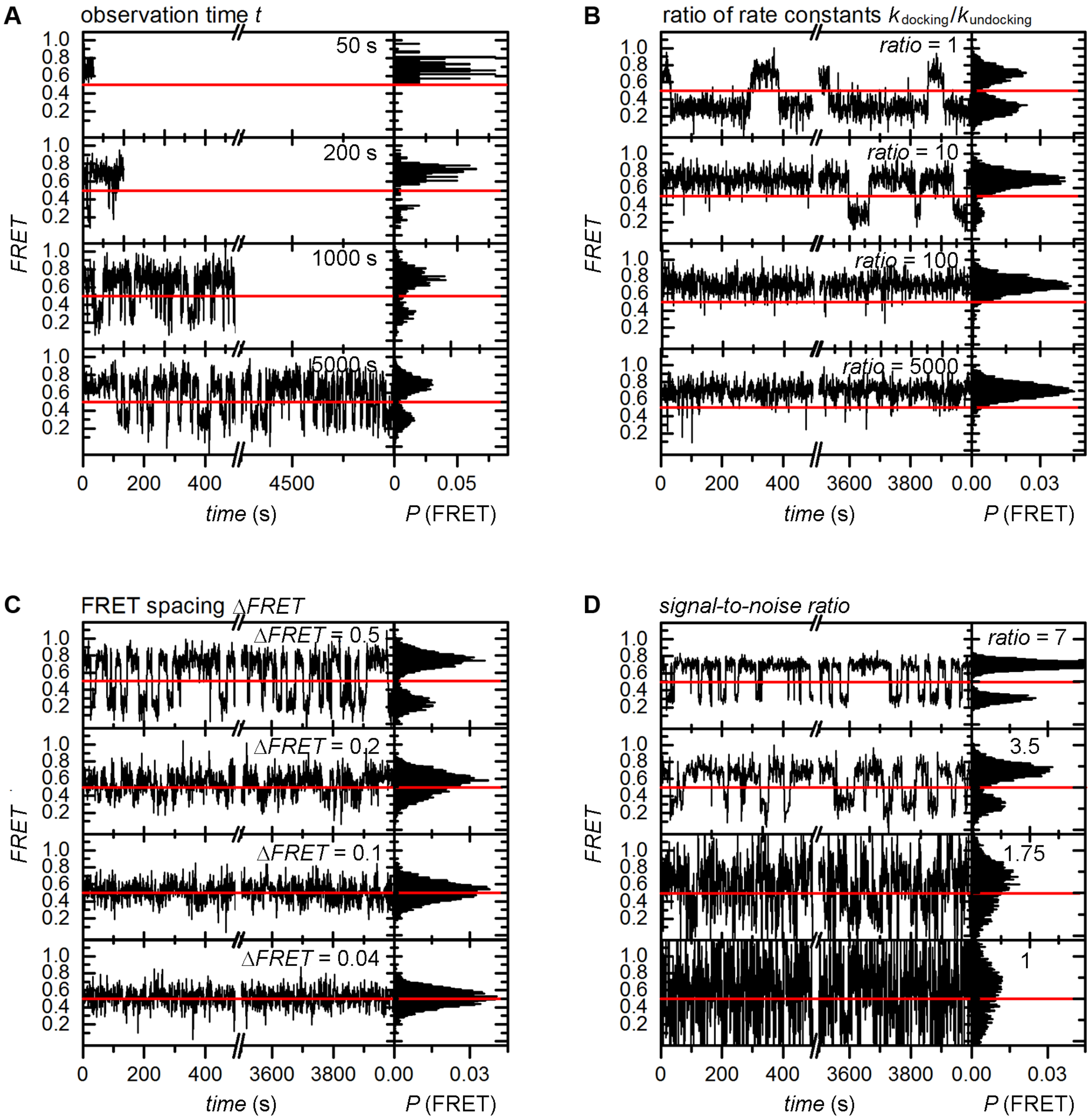


Figure 8.9: Representative data of simulated smFRET time traces and normalised histograms, representative data. Standard parameters of the simulation: $FRET_{\mathrm{A}}$ = 0.3 (undocked state), $FRET_{\mathrm{B}}$ = 0.7 (docked state); *SNR* = 3.5; observation time = 4000 s; $k_{\mathrm{docking}} = 0.1\ \mathrm{s}^{-1}$, $k_{\mathrm{undocking}} = 0.04\ \mathrm{s}^{-1}$. (A) The observation time *t* is varied between 50 s and 4000 s (1 frame per second). (B) The *ratio of rate constants* associated with docking and undocking is changed from 1 to 5000 ($k_{\mathrm{undocking}} = 0.005\ \mathrm{s}^{-1}$ = constant; $0.005\ \mathrm{s}^{11} \leq k_{\mathrm{docking}} \leq 25\ \mathrm{s}^{-1}$). (C) The spacing of the centers of the *FRET* distributions $\Delta FRET$ is varied from 0.5 to 0.02. (D) The *SNR* is varied from 7 to 1.

time-based methods perform poorly at short observation times, because the data before the first transition and preceding the last one are discarded. Importantly, the bootstrapped variability faithfully covers the theoretical values.

Figure 8.10B shows the influence of FRET spacing ($\Delta FRET$) on $\bar{X}_{\text{boba}}$ and $\sigma_{\text{boba}}$. In general, threshold-based approaches lead to a systematic downward shift of the estimated mean and estimations of cross-sample variability that do not cover the predicted values at low $\Delta FRET$ values. Similarly, HMM does not reliably distinguish the docked from the undocked state at $\Delta FRET < 0.1$. In turn, Gaussian fitting provides good estimations of the docked fraction, albeit $\sigma_{\text{boba}}$ is considerably more pronounced than for other methods at low $\Delta FRET$ values. The same trend is observed with decreasing *SNR* (Figure 8.10C). As $\Delta FRET$ and *SNR* diminish, the two *FRET* distributions get closer, becoming indistinguishable in extreme cases (Figure 8.9), explaining the bad performance of thresholding and why this approach should not be employed under these circumstances (Figure 8.1). HMM sets somewhat lower standards to the separation of the *FRET* distributions, though, it erroneously suggests equal population of both *FRET* states once it breaks down. Finally, even though the results of the Gaussian fits are biased by large error bars when the Rayleigh criterion is not fulfilled, the means are in excellent agreement with the theoretical values.

Figure 8.10D illustrates how the mean docked fraction and the cross-sample variability depend on the ratio of rate constants. Here, only the docking constant $k_{\text{docking}}$ is increased, while $k_{\text{undocking}}$ is kept constant at 0.005 $\text{s}^{-1}$, leading to a decreased average number of *FRET* transitions per time trace. As Gaussian fitting does not rely on faithful determination of dwell times, it provides an excellent estimation of the mean docked fraction and low cross-sample variability. In turn, threshold-based histogram analysis, HMM, and in particular threshold-based dwell time analysis underestimate the docked fraction when the thermodynamic equilibrium favors one conformation. Careful analysis of the *FRET* distributions from simulated smFRET time traces revealed that at *SNR* 3.5, the noise exceeds the threshold at times, explaining issues associated with thresholding. This is particularly problematic in the case of threshold-based dwell time analysis, as the ratio of false and true transitions then becomes highly unfavorable. In turn, when a conformational state is very scarcely populated, the mean dwell time becomes shorter than the time resolution and HMM fails to identify two FRET populations.

Figure 8.10E depicts the variation of $\bar{X}_{\text{boba}}$ and $\sigma_{\text{boba}}$ depending on the width of a *SNR* distribution, *i.e.* assuming intermolecular heterogeneity with regard to *SNR* within one dataset. For this purpose, *SNR* was assumed to be normally distributed around 3.5 and the width of the Gaussian distribution was varied between 0 (no heterogeneity) and 4 (strong heterogeneity). Analysis of *FRET* histograms and threshold-based dwell time analysis systematically underestimate the mean bound fraction by 3-5 %, which is due to the overlap between the two *FRET* states (*vide supra*). In turn, HMM-based dwell time analysis yields mean values and cross-sample variabilities that closely approach/cover the theoretical value in the case of narrow *SNR* distributions. However, as more low *SNR* time traces are included in the analysis, HMM perform increasingly

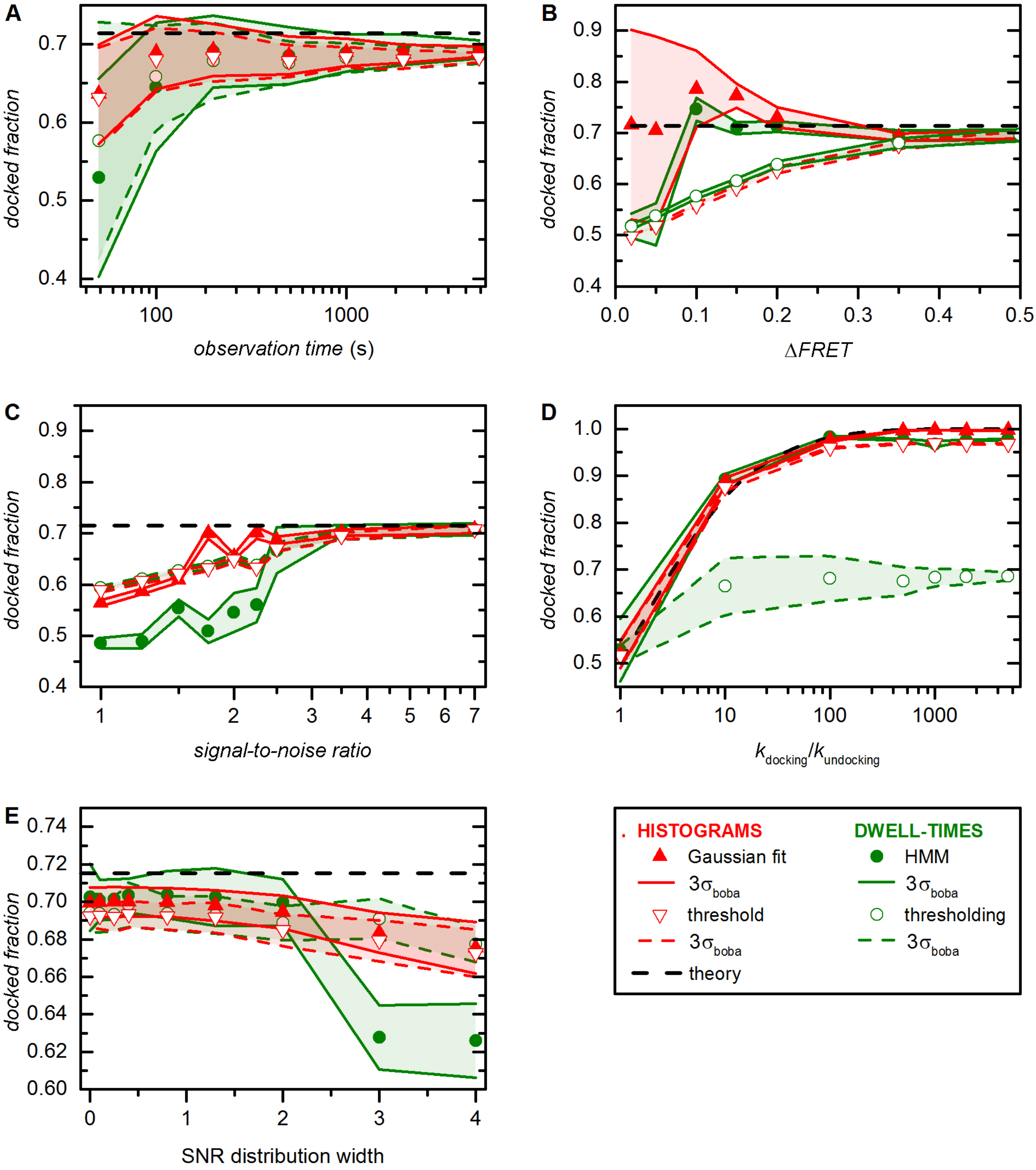


Figure 8.10: Robustness of different approaches for thermodynamic analysis of smFRET data performed in conjunction with bootstrapping (**method 1**, **2** and **3**, **thermodynamics**). Simulated data for a two-state system with standard parameters as defined in the methods section. (A) Performance in response to *trace length*. (B-C) Performance in response to *FRET spacing* and *SNR*. (D) Performance in response to the *ratio of rate constants*. (E) Performance in response to heterogeneously *distributed SNR* values. All theoretical values were determined from the input parameters used of the simulations. Error bars (red and green swaths) were estimated by bootstrapping and cover 99.7 % of the experimental variability ($3\sigma_{\text{boba}}$). Please refer to the text for further details.

poorly (*vide supra*). Interestingly, regardless of the method chosen for analysis, the estimation of the cross-sample variability remains mostly unaffected by a change in the width of the SNR distribution.

### 8.3.2 Kinetic characterisation of simulated smFRET data

When smFRET time traces display *discrete hops*, *i.e.* consist of piecewise constant signal, rate constants can be extracted from dwell time histograms (Figure 8.1) [444, 397]. Here, bootstrapping is applied to dwell times obtained by thresholding and HMM, followed by fitting the experimental data to a single-exponential decay model (both **method 1**, Equation 8.5, $O$ = 1) [394].

Figure 8.11A demonstrates that cross-sample variability strongly decreases when the observation time $t$ is increased from 50 s to 5000 s. Again, this is not surprising, as the average number of dwell times per time trace is expected to be proportional to the observation time, which leads to a more homogeneous behavior between individual time traces. However, thresholding systematically underestimates the mean decay constant associated with docking and undocking, an issue that noise is frequently mistaken as a *FRET* transition at *SNR* 3.5 (*vide supra*). This problem persists in HMM-based analysis, though, the algorithm proves more robust than thresholding.

The dependence of $\bar{X}_{\text{boba}}$ and $\sigma_{\text{boba}}$ on the number of dwell times is also depicted in Figure 8.11B, which shows the influence of the ratio of rate constants on the outcome of the dwell time analysis and the value of the bootstrapped error. Here, two effects lead to an underestimation of the decay constants: (i) Due to the (slight) overlap between the two *FRET* distributions at *SNR* = 3.5, there are short false dwell times stemming from noise (*vide supra*). As the ratio of $k_{\text{docking}}$ and $k_{\text{undocking}}$ increases the average number of true dwell times per time trace decreases, while the average number of noise-induced transitions is constant. Consequently, dwell times determined from individual time traces become more homogeneous as well, since false transitions are more and more emphasised. (ii) When the thermodynamic equilibrium strongly favors the docked state, brief excursions to the undocked state become irresolvable (*vide supra*).

Figures 8.11C and 8.11D show the influence of overlapping *FRET* distributions on the outcome of the kinetic analysis. In general, threshold-based analysis is strongly biased as the two *FRET* distributions display increasing overlap. As $\Delta$*FRET* and *SNR* diminish, the two *FRET* distributions display increasing overlap and the thresholding algorithm erroneously responds to noise, explaining its bad performance. Furthermore, cross-sample variability decreases, as each time trace (erroneously) yields a very high number of dwell times. The behavior of the HMM algorithm is not as easily explained: When the two *FRET* distributions display very strong overlap, HMM-based analysis yields approximately equal estimations for both for docking and undocking decay constants. Under such sub-Rayleigh conditions, the *FRET* distributions are essentially indistinguishable and the HMM algorithm (which assumes Gaussian noise) will ap-

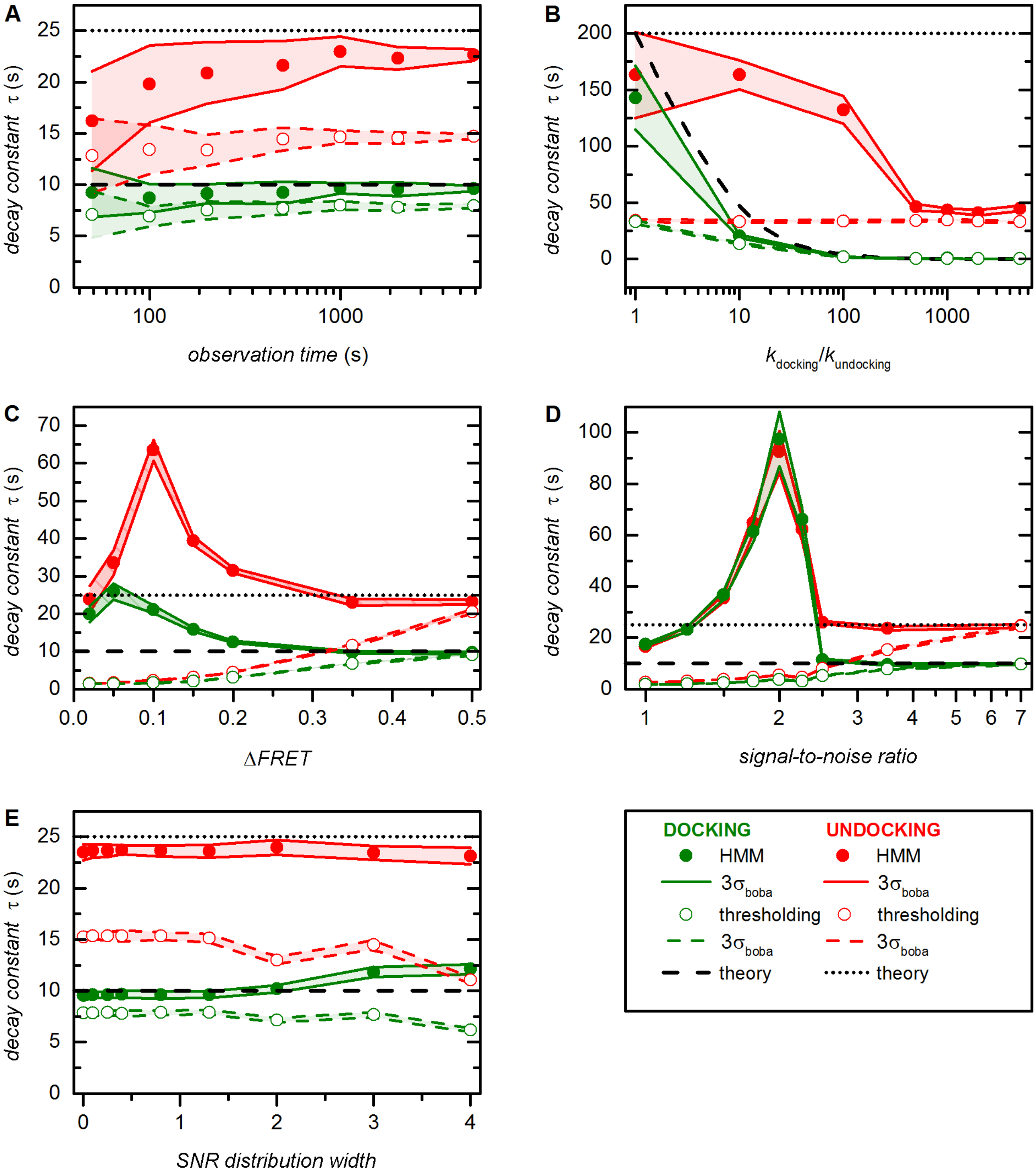


Figure 8.11: Robustness of thresholding and HMM approaches to analyse smFRET data performed in conjunction with bootstrapping (**method 1**, **kinetics**). Simulated data for a two-state system as defined in the methods section. (A) Performance in response to *trace length*. (B) Performance in response to the *ratio of rate constants*. (C-D) Performance in response to *FRET spacing* and *SNR*. (E) Performance in response to heterogeneously *distributed SNR*. All theoretical values were determined from the input parameters used of the simulations. Error bars (red and green swaths) correspond to the standard deviation estimated by bootstrapping ($3\sigma_{boba}$). Please refer to the text for further details.

proximate each time trace with two equally populated Gaussian distributions. As the two FRET distributions become more distinct, HMM tends to considerably overestimate the decay constants, which can be explained by (i) less artefactual transitions, and (ii) more real transitions. At the same time, however, not all transitions are identified, generally yielding an overestimation of the time a molecule dwells in the docked or undocked state. Further improvement of the data quality finally leads to a correct estimation at $\Delta FRET > 0.3$ and $SNR > 3.5$ and at $\Delta FRET > 0.4$ and $SNR > 2.5$. Importantly, throughout these simulations, the bootstrapped standard deviation is not significantly affected. In conclusion, HMM turns out to be more robust than thresholding in response to increasing overlap, an observation that is in excellent agreement with earlier reports [192].

Figure 8.11E shows how a variation of the *SNR* within the same dataset affects the estimation of $\bar{X}_{\text{boba}}$ and $\sigma_{\text{boba}}$. In general the influence of a change in *SNR distribution width* on both estimators is negligible. Threshold-based analysis consistently underestimates the values of the decay constants, which stems from the fact that the default *SNR* of 3.5 leads to a considerable number of erroneously identified dwell times as described above. In turn, the results of the HMM-based analysis are in good agreements with the theoretical prediction.

Taken together, these simulations illustrate the importance of selecting the correct method to analyse *FRET* time traces, as the bootstrapping algorithm cannot make up for ill-defined input values. However, when an appropriate approach is chosen, the bootstrapped cross-sample variability generally covers the theoretically predicted mean. Future work is anticipated to develop objective criteria to accept/reject a given model for thermodynamic and kinetic analysis of time-binned smFRET data presented herein.

## 8.4 Application of the algorithm to experimental data

Time-binned smFRET data have been recorded and analysed from numerous biological systems varying in size and complexity. Here, we studied an important element derived from the 5’ splice site recognition complex of the yeast group II intron *Sc*.ai5$\gamma$, the sequence pair d3’EBS1*/IBS1* [50, 388]. As depicted in Figure 8.2, Cy3- d3’EBS1* strands were tethered to the surface of a quartz slide passivated with biotinylated BSA, while Cy5-IBS1* molecules were free in solution. Hence, docking/undocking dynamics could be followed via FRET over several minutes and in the presence of different divalent metal ions, as splice site formation has previously been proposed to depend on the action of divalent metal ions [468]. FRET-typical anticorrelated changes in Cy3 and Cy5 emission intensity were observed in all cases, followed by calculating the *FRET* over time (Equation 8.1), which varied between zero (undocked) and a high *FRET* value (docked) for all dynamic molecules observed (Figure 8.12A). The fraction of statically undocked molecules, *i.e.* molecules that displays only donor emission during the time of observation, was 60 % in the absence of $M^{2+}$ and 20 % in the presence of $Ni^{2+}$ or $Co^{2+}$. This fraction of molecules either displays a low association constant $K_A$ that cannot be correctly resolved during the observation time and/or they correspond to a photophysical artifact, for example a docked IBS1* molecule with a non-emissive acceptor [146]. In fact, 15-55 % of the total population is usually *donor only* in smFRET studies using the FRET pair Cy3 and Cy5, which has been attributed to Cy5 pre-bleaching [469]. As a consequence, these

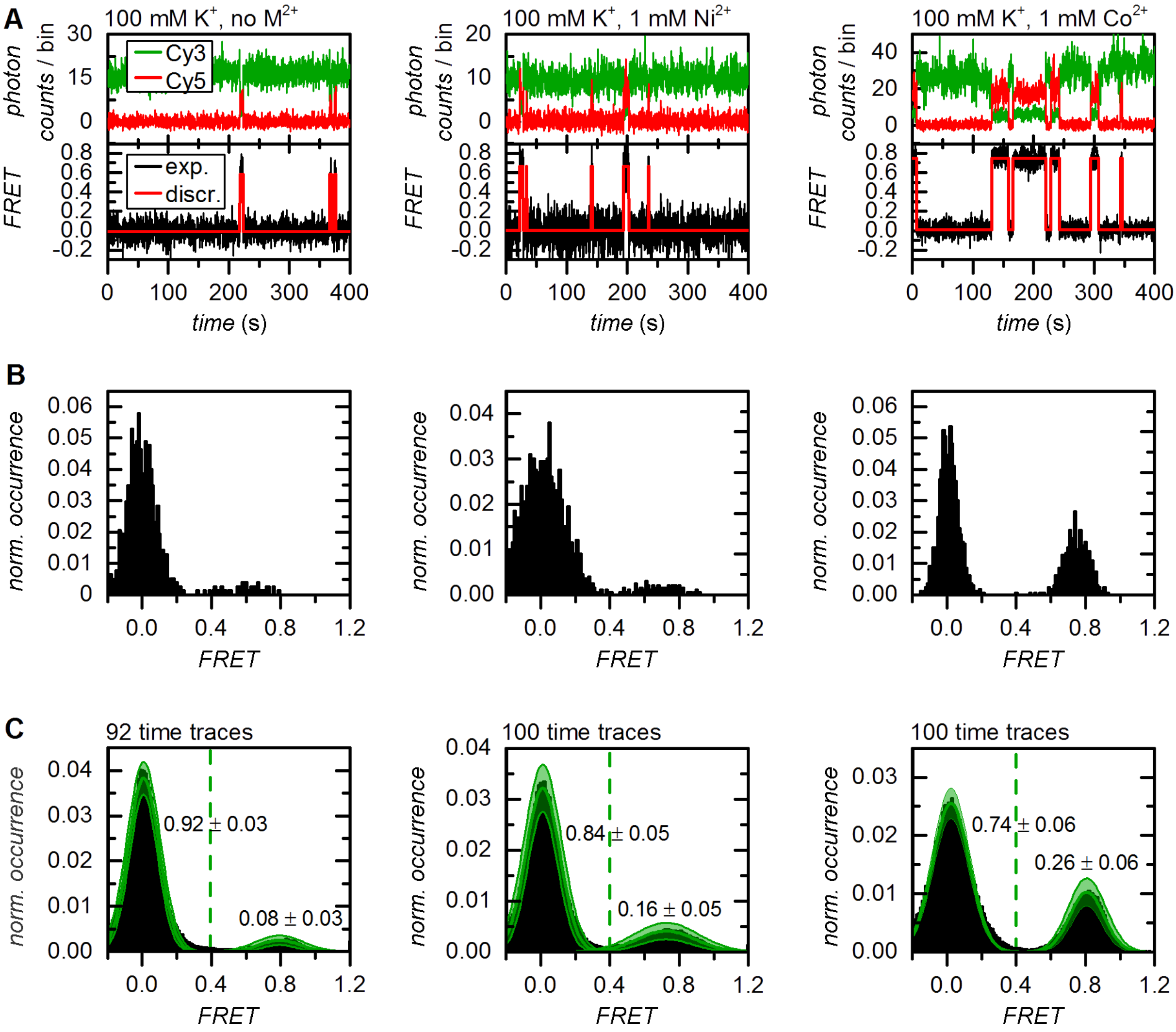


Figure 8.12: Representative time traces showing d3'EBS1*/IBS* interaction and thermodynamic analysis of *FRET* histograms (**method 1**). (A) Fluorophore emission over time reveals abrupt anticorrelated changes in *intensity* (upper graphs). Calculation of *FRET* time traces reveals repetitive shuttling between a zero and a high *FRET* level (lower graphs). Based on the experimental design, these two states were assigned to the undocked and the docked state. The red lines correspond to the discretisation by the Hidden Markov Model (vbFRET [394]). (B) *FRET* histograms built from the smFRET time traces shown in A. (C) Normalised cumulated *FRET* histograms built from individual time traces. The dashed green line depicts the threshold between the two *FRET* states used to determine the docked/undocked fractions and the normalised results are indicated. Solid green lines correspond to Gaussian approximation of the experimental data. The error (green swath) is the standard deviation associated with amplitude and width of the Gaussian fit functions as estimated by bootstrapping ($3\sigma_{\mathrm{boba}}$).

molecules were excluded from further analysis.

### 8.4.1 Divalent metal ions have a significant effect on the thermodynamic equilibrium

Bootstrapping was performed in conjunction with Gaussian fitting (**method 1**) and thresholding (**method 3**) of normalised cumulated *FRET* histograms (Figure 8.12B, C). The thermodynamic equilibrium was also characterised using dwell times obtained by HMM (**method 2**) [394]. Threshold-based analysis reveals weak inter-oligonucleotide interaction in the absence of divalent metal ions (docked fraction: 7.8 ± 2.6 %, errors correspond to $3\sigma_{\mathrm{boba}}$ unless specified differently, Figure 8.12C and Table 8.1). Addition of 1 mM $Ni^{2+}$ shifts the equilibrium slightly (docked fraction: 16.3 ± 3.3 %), while an average of 25.5 ± 6.0 % of all d3'EBS1* molecules are docked to IBS1* at 1 mM $Co^{2+}$. One-way analysis of variance (ANOVA) using bootstrapped values was performed to test the hypothesis that divalent metal ions affect or do not affect (null hypothesis) the thermodynamic equilibrium [429]. As illustrated in Figure 8.13A, an ANOVA makes the assumption that experimental values are normally distributed around the sample mean and its outcome (*P*-value) depends on the overlap integral between different distributions, which in turn depends on the separation of group means and the widths of the sample distributions. *P*-values constitute a strength of evidence against the null hypothesis and are typically compared to arbitrary values (0.05, 0.01 and 0.001) according to the conventions

Table 8.1: Thermodynamic analysis of the d3'EBS1*/IBS1* equilibrium by different methods. The experimental error was estimated by bootstrapping and accounts for 99.7 % of the variability observed ($3\sigma_{\mathrm{boba}}$, *68-95-99.7 rule* [429]). Association constants $K_A$ were calculated from normalised cumulated *FRET* histograms or dwell times under the assumption that $c(\mathrm{IBS^*}) = c_{\mathrm{tot}}(\mathrm{IBS1^*})$ (Equations 8.3 and 8.4).

| | imaging condition | fraction of docked d3'EBS1* | $K_A$ (L µmol$^{-1}$) |
|---|---|---|---|
| Gaussian fitting (**method 1**) | no $M(II)(NO_3)_2$ | 8.2 ± 3.0 | 3.6 ± 1.3 |
| | 1 mM $Ni(NO_3)_2$ | 18.5 ± 6.1 | 9.1 ± 3.7 |
| | 1 mM $Co(NO_3)_2$ | 28.6 ± 5.9 | 16.1 ± 4.6 |
| Dwell time analysis (HMM, **method 2**) | no $M(II)(NO_3)_2$ | 15.6 ± 6.4 | 7.4 ± 3.6 |
| | 1 mM $Ni(NO_3)_2$ | 24.6 ± 6.3 | 13.1 ± 4.5 |
| | 1 mM $Co(NO_3)_2$ | 33.3 ± 8.4 | 20.1 ± 7.5 |
| Histogram thresholding (**method 3**) | no $M(II)(NO_3)_2$ | 7.8 ± 2.6 | 3.4 ± 1.2 |
| | 1 mM $Ni(NO_3)_2$ | 16.3 ± 5.4 | 7.8 ± 3.1 |
| | 1 mM $Co(NO_3)_2$ | 25.5 ± 6.0 | 13.7 ± 4.5 |

of the field [429]. The presence of divalent metal ions not only significantly promotes the interaction of the two oligonucleotides ($P < 0.001$), the effect also differs significantly between $Ni^{2+}$ and $Co^{2+}$ ($P < 0.001$), the latter being much more effective in increasing the docked fraction (Figure 8.13A). Similar results were obtained by fitting the averaged 1D *FRET* histograms to two Gaussian distributions (Table 8.1), though the bootstrap-estimated errors are generally higher (Figure 8.12C). However, this did not strongly influence the significance of the effect ($P < 0.001$, data not shown). Thermodynamic analysis using dwell times (method 2) leads to a systematic shift of the mean docked fraction towards higher values and an increase of $\sigma_{\mathrm{boba}}$ (Table 8.1). Nevertheless, the results of all methods are generally in good agreement (Table 8.1).

Taken together, the results of histogram and dwell time analysis are in good agreement and demonstrate the significant role of low concentrations of divalent metal ions in shifting the thermodynamic equilibrium of d3'EBS1* and IBS1*. However, a systematic upward shift of the estimation of the docked fraction is observed that is most pronounced in the absence of divalent metal ions. These findings demonstrate that the dwell time approach has to be employed with care, especially when the biomolecule is poorly dynamic (60 % of statically undocked molecules in the absence of $M^{2+}$, *vide supra*) and/or the number of dwell times is

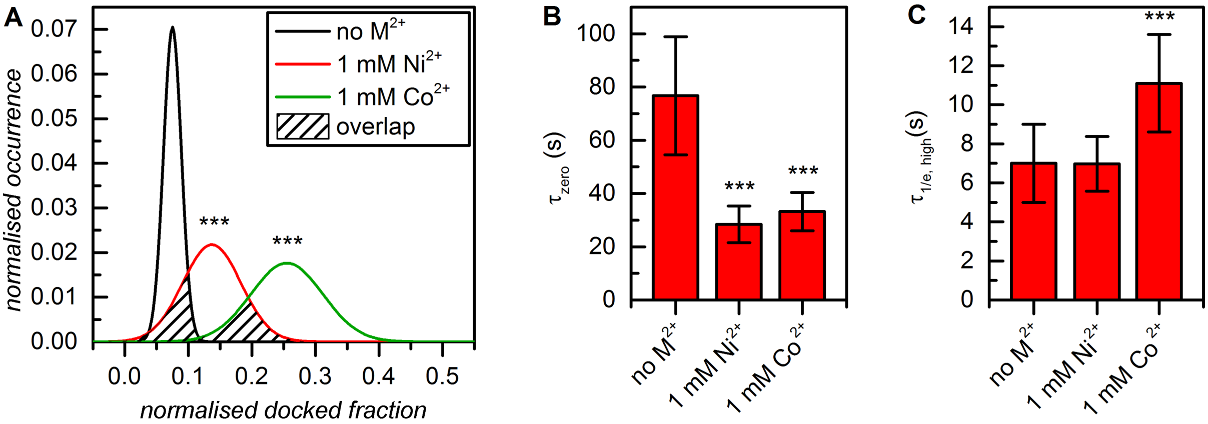


Figure 8.13: Statistical hypothesis testing using thermodynamic and kinetic smFRET data. (A) Analysis of variance (ANOVA) of docked fractions determined by thresholding of normalised cumulated *FRET* histograms (Figure 8.12C) reveals that $Ni^{2+}$ and $Co^{2+}$ shift the conformational equilibrium significantly towards the docked state (*** $P < 0.001$). The outcome of an ANOVA depends on the separation of the means (center values of the Gaussians) and how far the values are spread out (variance, $\sigma^2_{\mathrm{boba}}$, width of the Gaussians) and is given in form of a $P$-value, *i.e.* the probability that the null hypothesis is true (here: divalent metal ions do not significantly affect the equilibrium). The stronger the overlap between different groups, the greater the $P$-value. (B) Decay constants of the zero *FRET* state decrease in response to addition of $Ni^{2+}$ or $Co^{2+}$, leading to faster association ($P < 0.001$). Data obtained by HMM analysis and single-exponential fitting (Figure 8.14). (C) Decay constants of the high *FRET* state significantly increase in the presence of $Co^{2+}$ ($P < 0.001$), which promotes stable association of d3'EBS1* and IBS1*. Data obtained by HMM analysis and stretched exponential fitting. Error bars correspond to the bootstrapped standard deviation ($3\sigma_{\mathrm{boba}}$).

rather low, two problems that are often linked. Indeed, the average number of dwell times per time trace was less than 4, which contrast the average value of the simulations carried out using standard parameters (114, *vide supra*). As the first and the last dwell time were not considered, (i) much information was lost leading to an increase in the bootstrapped error and (ii) the occurrence of the more populated undocked state is underestimated translating into higher values of the docked fraction. Bias of dwell-time based approaches in the case of low numbers of dwell times can also be seen in the simulations (Figure 8.11A)

### 8.4.2 Divalent metal ions significantly alter d3'EBS1*/IBS1* interaction kinetics

d3'EBS1*/IBS1* dissociation has previously been shown to display considerable kinetic heterogeneity in the presence of divalent metal ions [164, 390, 470]. As a consequence, a stretched exponential decay (Equation 8.6) was fitted to dwell times in the high *FRET* state, while a single-exponential decay (Equation 8.5, $O = 1$) was used to approximate the association kinetics. Dwell times were determined from individual time traces using thresholding and HMM, followed by clustering of transition density plots using a weighted $k$-means algorithm (Figures 8.3 and 8.14A) [394, 432]. Then, cumulative probability plots *cumP* were created from dwell times, followed by fitting 1 - *normalised cumP* plots to exponential decay functions (Figure 8.14B-D). Dwell times were resampled via bootstrapping (**method 1**) to estimate the variability of the decay constants and the stretching parameter $\beta$.

HMM-based data on interstrand association is well described by single-exponential fit in the absence of divalent metal ions (Equation 8.6, $\beta = 0.99$ data not shown) and the process was found to occur very slowly ($\tau_{\text{docking}} = 76.7 \pm 22.2$ s). Both the presence of $Ni^{2+}$ and $Co^{2+}$ accelerates this reaction, albeit to different extents ($\tau_{\text{docking}} = 28.4 \pm 6.1$ s and $31.6 \pm 7.3$ s). These metal-ion-specific effects are highly significant as shown by one-way ANOVA ($P < 0.001$, Figure 8.13B). Importantly, the presence of divalent metal ions also induces slight broadening of the distribution of observed association rates ($\beta(Ni^{2+}) = 0.95$, $\beta(Co^{2+}) = 0.95$, data not shown), though the experimental data could nonetheless be satisfactorily approximated with the single-exponential fit (adjusted $R^2 > 0.98$ in all cases). d3'EBS1*/IBS1* dissociation is fast in the absence of divalent cations ($\tau_{\text{undocking, 1/e}} = 7.0 \pm 1.9$ s). $Co^{2+}$ significantly slows down the dissociation rate ($\tau_{\text{undocking}}$, 1/e $= 10.0 \pm 2.7$ s, $P < 0.001$), while the presence of $Ni^{2+}$ does not induce any variation in the decay constant ($\tau_{\text{undocking, 1/e}} = 7.0 \pm 1.4$ s, Figure 8.13C). In agreement with previous observations, the distributions of decay constants are severely broadened ($\beta < 0.9$ in all cases), underscoring the kinetic heterogeneity of the undocking process. The results of the threshold-based analysis are generally in excellent agreement with the values obtained from fitting HMM-derived dwell times. However, the decay constant associated with docking in the absence of divalent metal ions display a difference of 70 %. All results are summarised in Table 8.2.

These findings suggest that the presence of divalent metal ions broadens the distribution of rate constants associated with d3'EBS1*/IBS1* interaction. Based on the NMR structure and metal ion titration studies of the d3'EBS1* hairpin in the absence and presence of IBS1*, this effect has been assigned to heterogeneous occupation of metal ion binding sites along the RNA [390]. Such kinetic heterogeneity is beyond the scope of conventional kinetics and has frequently been observed in single-molecule experiments [163, 164]. In the context of this paper, kinetic heterogeneity contrasts the basic assumption made in first-order HMM, *i.e.* that state-to-state transitions are governed by single-exponential kinetics. The ability to assign one *FRET* level to multiple Markov transition rates is therefore important, an important feature that is implemented in some HMM software packages (vbFRET, CSSR) but not others (HaMMy) [192, 394, 437].

Fitting exponential decay models to bootstrapped dwell time histograms also permitted to show that changes in both association and dissociation kinetics are highly significant. Taken together, $Ni^{2+}$ shifts the thermodynamic equilibrium chiefly by promoting the association rate, while

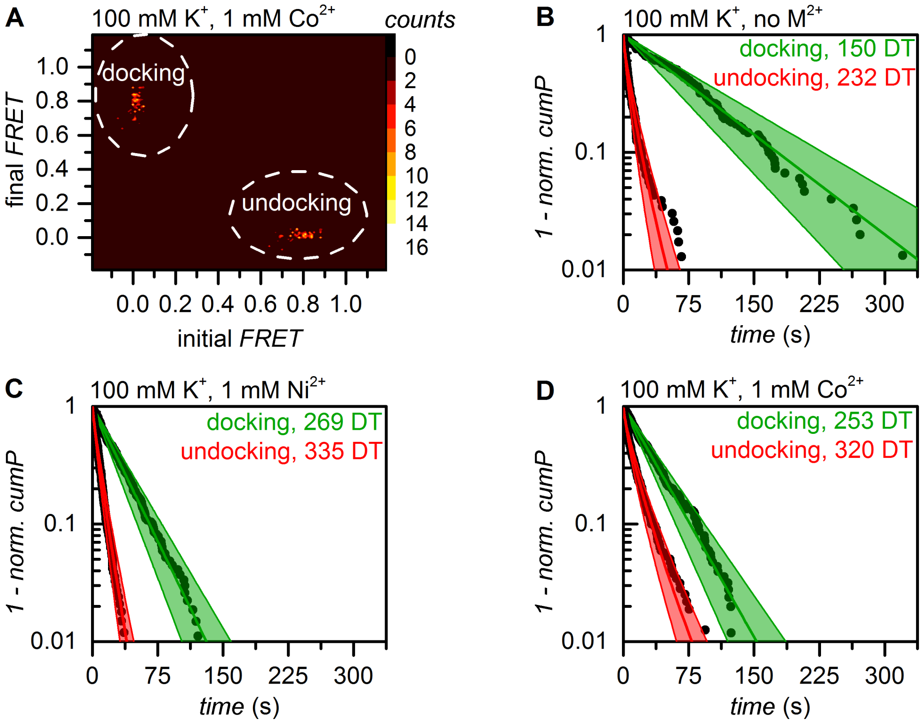


Figure 8.14: Kinetic analysis of smFRET data (method 1). Docking is defined as the state transition from the undocked to the docked state, the undocking process is defined as the inverse reaction. (A) Transition density plots of HMM data show two clusters corresponding to the docking and the undocking reaction, respectively. According to the maximum evidence approach employed in vbFRET [394], a two-state system is therefore most likely to produce the experimental data, which is in agreement with the experimental design. Raw data were grouped via the weighted *k*-means clustering algorithm. Colour code: occurrence in counts. (B-D) Dwell time histograms created from the normalised cumulative occurrence of dwell times in the docked and the undocked state as determined by HMM. The green lines correspond to a single-exponential fit to the experimental data, while the red lines represent a stretched exponential decay. Errors are indicated as a swath and correspond to $3\sigma_{\text{boba}}$ associated with the decay constants.

$Co^{2+}$ plays a two-fold role as an accelerator of docking and as an inhibitor of dissociation, probably by mediating specific contacts between the two RNA fragments. This difference is surprising, as both metal ions share very similar ionic radii ($Ni^{2+}$: 0.83 Å, $Co^{2+}$: 0.79 Å) and have the same preferred coordination geometry (octahedral, 6 ligands) [430]. Fits of threshold- and HMM-based dwell time data were generally in good agreement, except for docking in the absence of divalent cations. Careful analysis of HMM data revealed that brief excursions to the docked state were not always identified as such, especially when very few and short binding event occurred in the time trace (data not shown). Instead, the zero *FRET* distribution was erroneously identified as two distinct states. This observation contradicts the simulations and is most likely due to the fact that noise in experimental time traces does not always follow a stochastic Gaussian model (Figure 8.5). These findings suggests that HMM approaches are not always the best choice for analysing smFRET data, in particular when one conformation largely dominates the structural equilibrium and the occurrence of other structures may be erroneously deemed statistically insignificant by the HMM algorithm and non-Gaussian noise is fitted instead. As binding events became more frequent and/or long-lasting, HMM and thresholding were found to be in very good agreement.

Table 8.2: Kinetic analysis of d3'EBS1*/IBS1* association and dissociation using different methods to extract dwell times. The experimental error was estimated by bootstrapping and accounts for 99.7 % of the variability observed ($1\sigma_{\text{boba}}$ for $\beta$, $3\sigma_{\text{boba}}$ in all other cases). Rate constants were calculated according to Equations 8.7 and 8.8.

| | imaging condition | $\tau_{\text{dock}}{}^{a}$ (s) | $k_{\text{docking}}$ ($s^{-1}\mu M^{-1}$) | $\tau_{\text{undocking},1/e}{}^{b}$ (s) | $\beta$ | $k_{\text{undocking}}$ ($s^{-1}$) |
|---|---|---|---|---|---|---|
| Gaussian fitting (**method 1**) | no $M(II)(NO_3)_2$ | 44.9 ± 26.8 | 0.93 ± 0.60 | 7.6 ± 3.0 | 0.79 ± 0.08 | 0.156 ± 0.090 |
| | 1 mM $Ni(NO_3)_2$ | 31.7 ± 7.3 | 1.27 ± 0.29 | 6.2 ± 1.3 | 0.79 ± 0.04 | 0.175 ± 0.045 |
| | 1 mM $Co(NO_3)_2$ | 31.7 ± 7.9 | 1.27 ± 0.31 | 10.9 ± 2.6 | 0.75 ± 0.04 | 0.110 ± 0.033 |
| Dwell time analysis (HMM, **method 2**) | no $M(II)(NO_3)_2$ | 76.7 ± 22.2 | 0.53 ± 0.15 | 7.0 ± 2.0 | 0.77 ± 0.05 | 0.168 ± 0.057 |
| | 1 mM $Ni(NO_3)_2$ | 28.4 ± 6.1 | 1.41 ± 0.31 | 7.0 ± 1.4 | 0.88 ± 0.05 | 0.154 ± 0.042 |
| | 1 mM $Co(NO_3)_2$ | 33.2 ± 7.2 | 1.21 ± 70.26 | 11.1 ± 2.5 | 0.78 ± 0.04 | 0.105 ± 0.027 |

[a] single-exponential fit
[b] stretched exponential fit

## 8.5 Discussion

Single-molecule FRET has led to valuable work on mechanistic and structural aspects of numerous biological processes and has blossomed in recent years. However, the observation time of single fluorophore emission is rather limited, as dyes typically photobleach upon emission of $10^6$-$10^7$ photons (unpublished data involving Cy3 and Cy5 emission in the presence of an enzymatic oxygen scavenging system and 1 mM Trolox) [471]. Furthermore, the detected signal, intrinsically weak in intensity, is further broadened by various sources of additive noise and technical issues. As a consequence, single molecules typically display considerable cross-sample variability and can then not be treated as biological replicates in thermodynamic and kinetic analyses, *i.e.* rate and association constants cannot be inferred from individual smFRET time traces. In such cases, smFRET relies on the principle of ergodicity, according to which the properties of ensembles involving billions of molecules be described by combining a number of single molecules that is lower by several orders of magnitude [355]. Analogously, bootstrapping computes the distribution of the whole population, including measures of variance, from a sample distribution of the size $n$ [461].

Herschlag and co-workers have recently recognised the need for statistical rigor in smFRET experiments and implemented an HMM algorithm that assigned confidence intervals to rate constants inferred from individual time traces [397, 472]. Thus, one can investigate whether kinetically distinct subspecies exist within the sample, a long-standing topic of debate in the field of single-molecule spectroscopy [163, 164]. However, this approach sets very high standards to the data, as the confidence interval scales inversely to the number of transitions in the FRET time trace, and simulated time traces in the original article were composed of up to 5’000 dwell times [397]. Given the technical constraints outlined above, these values may be difficult to reach experimentally. Here, we have combined bootstrapping with different approaches commonly used in thermodynamic and kinetic analysis of smFRET data in order to estimate the variability associated with the mean values. By performing statistical hypothesis testing using generalised analysis of variance (ANOVA), we could show that divalent metal ions have a statistically significant effect on both thermodynamics and kinetics of d3’EBS1*/IBS1* interaction, a pair of RNA sequences involved in group II intron splice site recognition. Importantly, the fact that time traces were on average composed of only 4-6 dwell times was not problematic, since the overall data was treated as an ensemble according to the principle of ergodicity. We therefore believe that this approach is widely applicable and it is expected to make biological interpretations in smFRET experiments more robust when it is combined with statistical testing. Finally, it should be mentioned that the method described herein is not limited to time-binned smFRET data. We anticipate its implementation to analyse time traces stemming from single photon detection. A Further potential application is the characterisation of conformation and orientation dependent fluorophore photophysics (blinking, spectral and spatial diffusion) [395, 400, 473]. BOBA FRET was developed under Matlab version 8.20.701, license 49040 (Mathworks, Nattick, MA, USA) and is available at http://www.aci.uzh.ch/rna/.

# Chapter 9

# References

[1] Watson, J. D. and Crick, F. H. C. (1953) Molecular structure of nucleic acids, image downloaded from http://molbiochem.persiangig.com/. *Nature,* **171**, 737–738.

[2] Freisinger, E. and Sigel, R. K. O. (2007) From nucleotides to ribozymes - a comparison of their metal ion-binding properties. *Coord. Chem. Rev.,* **251**, 1834–1851.

[3] König, S. L. B., Evans, A. C., and Huppert, J. L. (2010) Seven essential questions on G-quadruplexes. *Biomol. Concepts,* **1**, 197–213.

[4] Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., and Goldfarb, A. (2003) Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J.,* **22**, 2234–2244.

[5] Beaucage, S. L., Iyer, , and  (1993) The synthesis of modified oligonucleotides by the phosphoramidite approach and their applications. *Tetrahedron,* **49**, 6123–6194.

[6] Deng, J., Xiong, Y., Pan, B., and Sundaralingam, M. (2003) Structure of an RNA dodecamer containing a fragment from SRP domain IV of *Escherichia coli*. *Acta Crystallogr., Sect. D,* **59**, 1004–1011.

[7] Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V., and Kopka, M. L. (1982) The anatomy of A-, B-, and Z-DNA. *Science,* **216**, 475–485.

[8] Varani, G. and McClain, W. H. (2000) The G-U wobble base pair. *EMBO Rep.,* **1**, 18–23.

[9] Aishima, J., Gitti, R. K., Noah, J. E., Gan, H. H., Schlick, T., and Wolberger, C. (2002) A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res.,* **30**, 5244–5252.

[10] Leontis, N. C. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA,* **7**, 499–512.

[11] Leontis, N. C. and Westhof, W. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.,* **30**, 3497–3531.

[12] Cisse, I. I., Kim, H., and Ha, T. (2012) A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.,* **19**, 623–627.

[13] Moore, P. B. The RNA folding problem. *In*: RNA worlds: from life's origins to diversity in gene regulation pp. 381–401 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA (1999).

[14] Hall, K., Cruz, P., Tinoco Jr., I., Jovin, T. M., and van de Sande, J. H. (1984) Z-RNA - a left-handed RNA double helix. *Nature,* **311**, 584–586.

[15] Kumari, S., Bugaut, A., Huppert, J. L., and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.,* **3**, 218–221.

[16] Huppert, J. L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.,* **37**, 1375–1384.

[17] König, S. L. B., Huppert, J. L., Sigel, R. K. O., and Evans, A. C. (2013) Distance-dependent duplex DNA destabilisation proximal to G-quadruplex/i-motif sequences. *Nucleic Acids Res.,* **41**, 7453–7461.

[18] McCallum, S. A. and Pardi, A. (2003) Refined structure of the iron-responsive element RNA using residual dipolar couplings. *J. Mol. Biol.,* **326**, 1037–1050.

[19] Sigel, R. K. O., Sashital, D. G., Abramovitz, D. L., Palmer, A. G., Butcher, S. E., and Pyle, A. M. (2004) The solution structure of domain 5 of a group II intron ribozyme reveals a new RNA motif. *Nat. Struct. Mol. Biol.,* **11**, 187–192.

[20] Khanna, M., Wu, H., Johansson, C., Caizergues-Ferrer, M., and Feigon, J. (2006) Structural study of the H/ACA snoRNP components Nop10p and the 3' hairpin of U65 snoRNA. *RNA,* **12**, 40–52.

[21] Donghi, D., Pechlaner, M., Finazzo, C., Knobloch, B., and Sigel, R. K. O. (2013) The structural stabilization of the $\kappa$ three-way junction by $Mg^{2+}$ represents the first step in the folding of a group II intron. *Nucleic Acids Res.,* **41**, 2489–2504.

[22] Brion, P. and Westhof, E. (1997) Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.,* **26**, 113–137.

[23] Batey, R. T., Rambo, R. P., and Doudna, J. A. (1999) Tertiary motifs in RNA structure and folding. *Angew. Chem. Int. Ed.,* **38**, 2326–2326.

[24] Treiber, D. K. and Williamson, J. R. (1999) Exposing the kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.,* **9**, 339–345.

[25] Buchmueller, K. L., Webb, A. E., Richardson, D. A., and Weeks, K. M. (2000) A collapsed non-native RNA folding state. *Nat. Struct. Biol.,* **7**, 362–366.

[26] Huang, Y., Zhang, J. L., Yu, X. L., Xu, T. S., Wang, Z. B., and Cheng, X. C. (2013) Molecular functions of small regulatory noncoding RNA. *Biochemistry (Moscow),* **78**, 221–230.

[27] Eddy, S. R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Gen.,* **2**, 919–929.

[28] König, S. L. B., Liyanage, P., Sigel, R. K. O., and Rueda, D. (2013) Helicase-mediated changes in RNA structure at the single-molecule level. *RNA Biol.,* **10**, 132–147.

[29] Schmeing, T. M. and Ramakrishnan, V. (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature,* **461**, 1234–1242.

[30] Physicky, E. M. and Hopper, A. K. (2010) tRNA biology charges to the front. *Gene Dev.,* **24**, 1832–1860.

[31] Faustino, N. A. and Cooper, T. A. (2003) Pre-mRNA splicing and human disease. *Gene Dev.,* **17**, 419–437.

[32] Bachellerie, J. P., Cavaillé, J., and Hüttenhofer, A. (2002) The expanding snoRNA world. *Biochimie,* **84**, 775–790.

[33] Ameres, S. L. and Zamore, P. D. (2013) Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell. Biol.,* **14**, 475–488.

[34] Nielsen, C. B., Shomron, N., Sandberg, R., Hornstein, E., Kitzman, J., and Burge, C. B. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA,* **13**, 1894–1910.

[35] Serganov, A. and Patel, D. J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Gen.,* **8**, 776–790.

[36] Crick, F. H. C. (1970) Central dogma of molecular biology. *Nature,* **227**, 561–563.

[37] Jacquier, A. (1996) Group II introns: elaborate ribozymes. *Biochimie,* **78**, 474–487.

[38] Doudna, J. A. and Cech, T. R. (2002) The chemical repertoire of natural ribozymes. *Nature,* **418**, 222–228.

[39] Roossinck, M. J., Sleat, D., and Palukaitis, P. (1992) Satellite RNAs of plant viruses: structures and biological effects. *Microbiol. Rev.,* **56**, 265–279.

[40] Kazantsev, A. V. and Pace, N. R. (2006) Bacterial RNase P: a new view of an ancient enzyme. *Nat. Rev. Microbiol.,* **4**, 729–740.

[41] Lambowitz, A. M., Caprara, M. G., Zimmerly, S., and Perlman, P. S. Group I and group II ribozymes as RNPs: clues to the past and guides to the future. *In*: RNA worlds: from life's origins to diversity in gene regulation pp. 451–485 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA (1999).

[42] Robart, A. R. and Zimmerly, S. (2005) Group II intron retroelements: function and diversity. *Cytogenet. Geneome Res.,* **110**, 589–597.

[43] Will, C. L. and Lührmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.,* **3**, a003707.

[44] Cordeaux, R. (2009) The impact of retrotransposons on human genome evolution. *Nat. Rev. Gen.,* **10**, 691–703.

[45] Lambowitz, A. M. and Zimmerly, S. (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.,* **3**, a003616.

[46] Li, W., Zhang, P., Fellers, J. P., Friebe, B., and Gill, B. S. (2004) Sequence composition, organization, and evolution of the core Triticeae genome. *Plant J.,* **40**, 500–511.

[47] Zhang, L. and Doudna, J. A. (2002) Structural insights into group II intron catalysis and branch-site selection. *Science,* **295**, 2084–2088.

[48] Erat, M. C. and Sigel, R. K. O. (2007) Determination of the intrinsic affinities of multiple site-specific $Mg^{2+}$ ions coordinated to domain 6 of a group II intron ribozyme. *Inorg. Chem.,* **46**, 11224–11234.

[49] Seetharam, M., Eldho, N. V., Padgett, R. A., and Dayie, K. T. (2006) Structure of a self-splicing group II intron catalytic effector domain 5: parallels with spliceosomal U6 RNA. *RNA,* **12**, 306–314.

[50] Kruschel, D Skilandat, M. and Sigel, R. K. O. (2014) NMR structure of the 5'-splice site in the group IIB intron Sc.aI5$\gamma$ - conformational requirements for positioning of the exon-intron junction. *RNA, in press,*.

[51] Pyle, A. M. (2010) The tertiary structure of group II introns: implications for biological function and evolution. *Crit. Rev. Biochem. Mol. Biol.,* **45**, 215–232.

[52] Vogel, J. and Borner, T. (2002) Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *EMBO J.,* **21**, 3794–3803.

[53] Li-Pook-Than, J. and Bonen, L. (2006) Multiple physical forms of excised group II intron RNAs in wheat mitochondria. *Nucleic Acids Res.,* **34**, 2782–2790.

[54] Molina-Sanchez, M. D., Martinez-Abaca, F., and Toro, N. (2006) Excision of the *Sinorhizobium meliloti* group II intron RmInt1 as circles *in vivo*. *J. Biol. Chem.,* **281**, 28737–28744.

[55] Pyle, A. M. and Lambowitz, A. M. Group II introns: ribozymes that splice RNA and invade DNA. *In*: The RNA world pp. 469–505 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA (2006).

[56] Somarowthu, S., Legiewicz, M., Keating, K. S., and Pyle, A. M. (2013) Visualizing the ai5$\gamma$ group IIB intron. *Nucleic Acids Res., in press,*.

[57] Fedorova, O. and Zingler, N. (2007) Group II introns: structure, folding and splicing mechanism. *Biol. Chem.,* **388**, 665–678.

[58] Michel, F. and Ferat, J. L. (1995) Structure and activities of group II introns. *Annu. Rev. Biochem.,* **64**, 435–461.

[59] Fedorova, O., Mitros, T., and Pyle, A. M. (2003) Domains 2 and 3 interact to form critical elements of the group II intron active site. *J. Mol. Biol.,* **330**, 197–209.

[60] Lambowitz, A. M. and Zimmerly, S. (2004) Mobile group II introns. *Annu. Rev. Genet.,* **38**, 1–35.

[61] Boudvillain, M. and Pyle, A. M. (1998) Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.,* **17**, 7091–7104.

[62] Boudvillain, M., de Lencastre, A., and Pyle, A. M. (2000) A tertiary interaction that links active-site domains to the 5' splice site of a group II intron. *Nature,* **406**, 315–318.

[63] Schnabl, J., Sutter, P., and Sigel, R. K. O. (2011) MINAS - a database of Metal Ions in Nucleic AcidS. *Nucleic Acids Res.,* **40**, D434–D438.

[64] Sigel, H. and Sigel, R. K. O. Metal ion interactions with nucleic acids and their constituents. *In*: Comprehensive Inorganic Chemistry II pp. 623–660 Elsevier Ltd., Oxford, United Kingdom (2013).

[65] Auffinger, P., Grover, N., and Westhof, E. (2011) Metal ion binding to RNA. *Met. Ions Life Sci.,* **9**, 1–35.

[66] Pechlaner, M. and Sigel, R. K. O. (2012) Interplay between metal ions and nucleic acids. *Met. Ions Life Sci.,* **9**, 1–42.

[67] Miskin, R., Zamir, A., and Elson, D. (1970) Inactivation and reactivation of ribosomal subunits: the peptidyl transferase activity of the 50 s subunit of *Escherichia coli*. *J. Mol. Biol.,* **54**, 355–366.

[68] Zamir, A., Miskin, R., and Elson, D. (1971) Inactivation and reactivation of ribosomal subunits: Amino acyl-transfer RNA binding activity of the 30 s subunit of *Escherichia coli*. *J. Mol. Biol.,* **60**, 347–352.

[69] Cahn, F. and Lubin, M. (1978) Inhibition of elongation steps of protein synthesis at reduced potassium concentrations in reticulocytes and reticulocyte lysate.. *J. Biol. Chem.,* **253**, 7798–7803.

[70] Näslund, P. H. and Hultin, T. (1970) Effects of potassium deficiency on mammalian ribosomes. *Biochim. Biophys. Acta,* **204**, 237–247.

[71] Näslund, P. H. and Hultin, T. (1971) Structural and functional defects in mammalian ribosomes after potassium deficiency. *Biochim. Biophys. Acta,* **254**, 104–116.

[72] Ennis, H. L. and Artman, M. (1972) Ribosome size distribution in extracts of potassium-depleted Escherichia coli. *Biochem. Biophys. Res. Commun.,* **48**, 161–168.

[73] Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S., and Kourreck, J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA,* **14**, 1290–1296.

[74] Bugaut, A. and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: tranlation regulation and targeting. *Nucleic Acids Res.,* **40**, 4727–4741.

[75] Martadinata, H. and Phan, A. T. (2013) Structure of human telomeric RNA (TERRA): stacking of two G-quadruplex blocks in K(+) solution. *Biochemistry,* **52**, 2176–2183.

[76] Sigel, R. K. O., Vaidya, A., and Pyle, A. M. (2000) Metal ion binding sites in a group II intron core. *Nat. Struct. Biol.,* **7**, 1111–1117.

[77] Erat, M. C., Zerbe, O., Fox, T., and Sigel, R. K. O. (2007) Solution structure of domain 6 from a self-splicing group II intron ribozyme: a $Mg^{2+}$ binding site is located close to the stacked branch adenosine. *Chem. Bio. Chem.,* **8**, 306–314.

[78] Erat, M. C. and Sigel, R. K. O. (2008) Divalent metal ions tune the self-splicing reaction of the yeast mitochondrial group II intron Sc.ai5$\gamma$. *J. Biol. Inorg. Chem.,* **13**, 1025–1036.

[79] Waldsich, C. and Pyle, A. M. (2007) A folding control element for tertiary collapse of a group II ribozyme. *Nat. Struct. Biol.,* **14**, 37–44.

[80] Waldsich, C. and Pyle, A. M. (2008) A kinetic intermediate that regulates proper folding of a group II intron RNA. *J. Mol. Biol.,* **375**, 572–580.

[81] Klein, D. J., Moore, P. B., and Steitz, A. (2004) The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA,* **10**, 1366–1379.

[82] Hsiao, C. and Williams, L. D. (2009) A recurrent magnesium-binding motif provides a framework for the ribosomal peptidyl transferase center. *Nucleic Acids Res.,* **37**, 3134–3142.

[83] Cromie, M. J., Shi, Y., Latifi, T., and Groisman, E. A. (2006) An RNA sensor for intracellular Mg(2+). *Cell,* **125**, 71–84.

[84] Brantl, S. (2006) Bacterial gene regulation: Metal ion sensing by proteins or RNA. *Trends Biotechnol.,* **24**, 383–386.

[85] Dann, 3rd, C. E., Wakeman, C. A., Siedling, C. L., Baker, S. C., Irnov, I., and Winkler, W. C. (2007) Structure and mechanism of a metal-sensing regulatory RNA. *Cell,* **130**, 878–892.

[86] Wakeman, C. A., Ramesh, S., and Winkler, W. C. (2009) Multiple metal binding cores are required for metalloregulation by M box riboswitch RNAs. *J. Mol. Biol.,* **392**, 723–735.

[87] Steiner, M., Rueda, D., and Sigel, R. K. O. (2009) $Ca^{2+}$ induces the formation of two distinct subpopulations of group II intron molecules. *Angew. Chem. Int. Ed.,* **48**, 9739–9742.

[88] Seyhan, A. A., Amaral, J., and Burke, J. M. (1998) Intracellular RNA cleavage by the hairpin ribozyme. *Nucleic Acids Res.,* **26**, 3494–3504.

[89] Oh, B. K., Frank, D. N., and Pace, N. R. (1998) Participation of the 3'-CCA of tRNA in the binding of catalytic $Mg^{2+}$ ions by ribonuclease P. *Biochemistry,* **37**, 7277–7283.

[90] Ke, A., Zhou, K., Ding, F., Cate, J. H., and Doudna, J. A. (2004) A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature,* **429**, 201–205.

[91] Roychowdhury-Saha, M. and Burke, D. H. (2006) Extraordinary rates of transition metal ion-mediated ribozyme catalysis. *RNA,* **14**, 1846–1852.

[92] Khan, M. A., Walden, W. E., Goss, D. J., and Theil, E. C. (2009) Direct $Fe^{2+}$ sensing by iron-responsive messenger RNA-repressor complexes weakens binding. *J. Biol. Chem.,* **284**, 30122–30128.

[93] Hsiao, C., Chou, I. C., Okafor, C. D., Bowman, J. C., O'Neill, E. B., Athavale, S. S., Petrov, A. S., Hud, N. V., Wartell, R. M., Harvey, S. C., and Williams, L. D. (2013) RNA with iron(II) as a cofactor catalyses electron transfer. *Nat. Chem.,* **5**, 525–528.

[94] Nahvi, A., Barrick, J. E., and Breaker, R. R. (2004) Coenzyme $B_{12}$ riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.,* **32**, 143–150.

[95] Gallo, S., Oberhuber, M., Sigel, R. K. O., and Kräutler, B. (2008) The corrin moiety of coenzyme $B_{12}$ is the determinant for switching the *btub* riboswitch of *E. coli*. *ChemBioChem,* **9**, 1408–1414.

[96] Borda, E. J., Markley, J. C., and Sigurdsson, S. T. (2004) Zinc-dependent cleavage in the catalytic core of the hammerhead ribozyme: evidence for a pH-dependent conformational change. *Nucleic Acids Res.,* **31**, 2595–2600.

[97] Heilman-Miller, S. L., Thirumalai, D., and Woodson, S. (2001) Role of counterion condensation in folding of the *Tetrahymena* ribozyme. I. Equilibrium stabilization by cations. *J. Mol. Biol.,* **306**, 1157–1166.

[98] Cowan, J. A. (1995) The biological chemistry of magnesium, VCH Publishers, Inc., New York City, NY, USA.

[99] Shiman, R. and Draper, D. E. (2000) Stabilization of RNA tertiary structure by monovalent cations. *J. Mol. Biol.,* **302**, 79–91.

[100] Crichton, R. R. (2008) Biological inorganic chemistry, Elsevier B.V., Amsterdam, The Netherlands.

[101] Boots, J. L., Canny, M. D., Azimi, E., and Pardi, A. (2008) Metal ion specificities for folding and cleavage activity in the *Schistosoma* hammerhead ribozyme. *RNA,* **14**, 2212–2222.

[102] Sigel, R. K. O. (2005) Group II intron ribozymes and metal ions - a delicate Relationship. *Eur. J. Inorg. Chem.,* **12**, 2281–2292.

[103] Osborne, E. M., Schaak, J. E., and DeRose, V. J. (2005) Characterization of a native hammerhead ribozyme derived from schistosomes. *RNA,* **11**, 187–196.

[104] Burton, A. S. and Lehman, N. (2006) Calcium(II)-dependent catalytic activity of the *Azoarcus* ribozyme: testing the limits of resolution for in vitro selection. *Biochemie,* **88**, 819–825.

[105] Chu, V. B., Bai, Y., Lipfert, J., Herschlag, D., and Doniach, S. (2008) A repulsive field: advances in the electrostatics of the ion atmosphere. *Curr. Opin. Chem. Biol.,* **12**, 619–625.

[106] Draper, D. E., Grilley, D., and Soto, A. M. (2005) Ions and RNA folding. *Annu Rev. Biophys. Biomol. Struct.,* **34**, 221–243.

[107] Schnabl, J. and Sigel, R. K. O. (2010) Controlling ribozyme activity by metal ions. *Curr. Op. Chem. Biol.,* **14**, 269–275.

[108] Kasha, M. (1950) Characterization of electronic transitions in complex molecules. *Disc. Faraday Soc.,* **9**, 14–19.

[109] Lakowicz, J. R. (2006) Principles of fluorescence spectroscopy, Springer Science+Business Media, LCC, New York City, NY, USA, .

[110] Stokes, G. G. (1852) On the change of refrangibility of light. *Phil. Trans. R. Soc.,* **142**, 463–562.

[111] Ha, T. (2001) Single-molecule fluorescence energy transfer. *Methods,* **25**, 78–86.

[112] Laverdant, J., de Marcillac, W. D., Barthou, C., Chinh, V. D., Schwob, C., Coolen, L., Benalloul, P., Nga, P. T., and MaÃ®tre, A. (2011) Experimental determination of the fluorescence quantum yield of semiconductor nanocrystals. *Materials,* **4**, 1182–1193.

[113] Berezin, M. Y. and Achilefu, S. (2010) Fluorescence lifetime measurements and biological imaging. *Chem. Rev.,* **110**, 2641–2684.

[114] Dexter, D. L. (1953) A theory of sensitized luminescence in solids. *J. Chem. Phys.,* **21**, 836–850.

[115] Varnes, A. W., Dodson, R. B., and Wehry, E. L. (1972) Interactions of transition-metal ions with photoexcited states of flavins. Fluorescence quenching studies. *J. Am. Chem. Soc.,* **94**, 946–950.

[116] Kautsky, H. (1939) Quenching of luminescence by oxygen. *Trans. Faraday Soc.,* **35**, 216–219.

[117] Knibbe, H., Rehm, D., and Weller, A. (1968) Intermediates and kinetics of fluorescence quenching by electron transfer. *Ber. Bunsenges. Phys. Chem.,* **72**, 257–263.

[118] Birch, D. J. S., Suhling, K., Holmes, A. S., Salthammer, T., and Imhof, R. E. (1993) Metal ion quenching of perylene fluorescence in lipid bilayers. *Pure & Appl. Chem.,* **65**, 1687–1692.

[119] Cordes, T., Santoso, Y., Tomescu, A. I., Gryte, K., Hwang, L. C., Camara, B., Wigneshweraraj, S., and Kapanidis, A. N. (2010) Sensing DNA opening in transcription using quenchable Förster resonance energy transfer. *Biochemistry,* **49**, 9171–9180.

[120] Pustovit, V. N. and Shahbazyan, T. V. (2012) Fluorescence quenching near small metal nanoparticles. *J. Chem. Phys.,* **136**, 204701.

[121] Dahan, M., Deniz, A. A., Ha, T., Chemla, D. S., Schultz, P. G., and Weiss, S. (1999) Ratiometric measurement and identification of single diffusing molecules. *Chem. Phys.,* **247**, 85–106.

[122] Rasnik, I., McKinney, S. A., and Ha, T. (2005) Surfaces and orientations: much to FRET about. *Acc. Chem. Res.,* **38**, 542–548.

[123] Kahra, D., Kovermann, M., Löw, C., Hirschfeld, V., Haupt, C., Balbach, J., and Hübner, C. G. (2011) Conformational plasticity and dynamics in the generic protein folding catalyst SlyD unraveled by single-molecule FRET. *J. Mol. Biol.,* **411**, 781–790.

[124] Hohlbein, J., Gryte, K., Heilemann, M., and Kapanidis, A. N. (2010) Surfing on a new wave of single-molecule fluorescence methods. *Phys. Biol.,* **7**, 031001.

[125] Lee, S., Lee, J., and Hohng, S. (2010) Single-molecule three-color FRET with both negligible spectral overlap and long observation time. *PLoS ONE,* **5**, e12270.

[126] Stryer, L. and Haugland, R. P. (1967) Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. USA,* **58**, 719–726.

[127] Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M., and Eaton, W. A. (2005) Polyproline and the “spectroscopic ruler” revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA,* **102**, 2754–2759.

[128] Vos, W. L., Koehorst, R. B. M., Spruijt, R. B., and Hemminga, M. A. (2005) Membrane-bound conformation of M13 major coat protein. *J. Biol. Chem.,* **280**, 38522–38527.

[129] Iqbal, A., Arshlan, S., Okumus, B., Wilson, T. J., Giraud, G., Norman, D. G., Ha, T., and Lilley, D. M. J. (2008) Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl. Acad. Sci. USA,* **105**, 11176–11181.

[130] Iqbal, A., Wang, L., Thompson, K. C., Lilley, D. M. J., and Norman, D. G. (2008) The structure of cyanine 5 terminally attached to double-stranded DNA: implications for FRET studies. *Biochemistry,* **47**, 7857–7862.

[131] Sindbert, S., Kalinin, S., Nguyen, H., Kienzler, A., Clima, L., Bannwarth, W., Appel, B., Müller, S., and Seidel, C. A. M. (2011) Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *J. Am. Chem. Soc.,* **133**, 2463–2480.

[132] Owicki, J. C. (2000) Fluorescence polarization and anisotropy in high throughput screening: perspectives and primer. *J. Biomol. Screen.,* **5**, 297–306.

[133] Ameloot, M., ven de Ven, M., Acuna, A. U., and Valeur, B. (2013) Fluorescence anisotropy measurements in solution: methods and reference materials. *Pure Appl. Chem.,* **85**, 589–608.

[134] Valeur, B. (2001) Molecular fluorescence: principles and applications, Wiley-VCH, New York City, NY, USA, .

[135] Vu, H. M., Pasternack, L. B., and Kearns, D. R. (1999) Specificity of hydroxylmethyluracil-containing DNA for transcription factor 1: structural insights. *,* **52**, 57–63.

[136] Vljayendran, R. A. and Leckband, D. E. (2001) A quantitative assessment of heterogeneity for surface-immobilized proteins. *Anal. Chem.,* **73**, 471–480.

[137] Borst, J. W., Hink, M. A., van Hoek, A., and Visser, A. J. (2005) Effects of refractive index and viscosity on fluorescence and anisotropy decays of enhanced cyan and yellow fluorescent proteins. *J. Fluoresc.,* **15**, 153–160.

[138] Wiersma, D. S., Muzzi, A., Muzzi, A., Colocci, M., and Ringhini, R. (1999) Time-resolved anisotropic multiple light scattering in nematic liquid crystals. *Phys. Rev. Lett.,* **83**, 4321–4324.

[139] Sapsford, K. E., Berti, L., and Medintz, I. L. (2006) Materials for fluorescence resonance energy transfer analysis: beyond traditional donor-acceptor combinations. *Angew. Chem. Int. Ed.,* **45**, 4562–4589.

[140] Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science,* **273**, 1392–1395.

[141] Roy, R., Hohng, S., and Ha, T. (2008) A practical guide to single-molecule FRET. *Nat. Methods,* **5**, 507–516.

[142] Darby, R. A. J., Sollogoub, M., McKeen, C., Brown, L., Risitano, A., Brown, N., Barton, C., and Fox, K. R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Res.,* **30**, e39.

[143] Shimomura, O., Johnson, F. H., and Saiga, Y. (1962) Extraction, purification and properties of Aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J. Cell Comp. Physiol.,* **59**, 223–239.

[144] Shaner, N., Steinbach, P., and Tsien, R. Y. (2005) A guide to choosing fluorescent proteins. *Nat. Methods,* **2**, 905–915.

[145] Hohng, S. and Ha, T. (2005) Single-molecule quantum-dot fluorescence resonance energy transfer. *ChemPhysChem,* **6**, 956–960.

[146] Ha, T. and Tinnefeld, P. (2012) Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.,* **63**, 595–617.

[147] Martin, M. M. and Lindquist, L. (1975) The pH dependence of fluorescein fluorescence. *J. Lumin.,* **10**, 381–390.

[148] Shi, X., Lim, J., and Ha, T. (2010) Acidification of the oxygen scavenging system in single-molecule fluorescence studies: in situ sensing with a ratiometric dual-emission probe. *Anal. Chem.,* **82**, 6132–6138.

[149] Aitken, C. E., Marshall, R. A., and Puglisi, J. D. (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.,* **94**, 1826–1835.

[150] Rasnik, I., McKinney, S. A., and Ha, T. (2006) Nonblinking and longlasting single-molecule fluorescence imaging. *Nat. Methods,* **3**, 891–893.

[151] Cordes, T., Vogelsang, J., and Tinnefeld, P. (2009) On the mechanism of Trolox as antiblinking and antibleaching agent. *J. Am. Chem. Soc.,* **131**, 5018–5019.

[152] Huang, Q. and Fu, W. L. (2005) Comparative analysis of the DNA staining efficiencies of different fluorescent dyes in preparative agarose gel electrophoresis. *Clin. Chem. Lab. Med.,* **43**, 841–842.

[153] Sanger, F. and Coulson, A. R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.,* **94**, 441–448.

[154] Guo, J., Xu, N., Li, Z., Zhang, S., Wu, J., Kim, D. H., Marma, M. S., Meng, Q., Cao, H., Li, X., Shi, S., Yu, L., Kalachikov, S., Russo, J. J., Turro, N. J., and Ju, J. (2008) Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl. Acad. Sci. USA,* **105**, 9145–9150.

[155] VanGuilder, H. D., Vrana, K. E., and Freeman, W. M. (2008) Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques,* **44**, 619–626.

[156] Jackson Immuno Research Laboratories, inc. (Oct 2013) Technical information on probes conjugated to affinity-purified antibodies and to other proteins, http://www.jacksonimmuno.com/technical/f-cy3-5.asp.

[157] Gräff, J., Woldemichael, B. T., Berchtold, D., Dewarrat, G., and Mansuy, I. M. (2012) Dynamic histone marks in the hippocampus and cortex facilitate memory consolidation. *Nat. Commun.,* **3**, 991.

[158] Fulwyler, M. J. (1965) Electronic separation of biological cells by volume. *Science,* **150**, 910–911.

[159] Langer-Safer, P. R., Levine, M., and Ward, D. C. (1982) Immunological method for mapping genes on *Drosophila* polytene chromosomes. *Proc. Natl. Acad. Sci. USA,* **79**, 4381–4385.

[160] Philips, G. J. (2001) Green fluorescent protein - a bright idea for the study of bacterial protein localization. *FEMS Microbiol. Lett.,* **16**, 9–18.

[161] Walter, N. G., Huang, C. Y., Manzo, A. J., and Sobhy, M. A. (2008) Do-it-yourself guide: how to use the modern single-molecule toolkit. *Nat. Methods,* **5**, 475–489.

[162] Feynman, R. P. Miniaturization pp. 282–296 Reinhold Publishing Cooperation, New York City, NY, USA (1961).

[163] Zhao, R. and Rueda, D. Memory effects in RNA folding dynamics revealed by single-molecule fluorescence. *In*: Biophysics of RNA folding (Biophysics for the Life Sciences) pp. 117–133 Springer, New York City, NY, USA (2012).

[164] König, S. L. B., Kowerko, D., and Sigel, R. K. O. (2013) Kinetic subpopulations detected by single-molecule spectroscopy: fundamental property of functional nucleic acids or experimental artefact. *CHIMIA,* **67**, 240–243.

[165] Shirude, P. S. and Nalasubramanian, S. (2008) Single-molecule conformational analysis of DNA G-quadruplexes. *Biochimie,* **90**, 1197–1206.

[166] König, S. L. B., Hadzic, M., Fiorini, E., Börner, R., Kowerko, D., Blanckenhorn, W. U., and Sigel, R. K. O. (2013) BOBA FRET: Bootstrap-based analysis of single-molecule FRET data. *PLoS ONE,* **8**, e84157.

[167] Huang, B., Bates, M., and Zhuang, X. (2009) Super resolution fluorescence microscopy. *Annu. Rev. Biochem.,* **78**, 993–1016.

[168] Honda, M., Park, J., Pugh, R. A., Ha, T., and Spies, M. (2009) Single-molecule analysis reveals differential effect of ssDNA-binding proteins on DNA translocation by XPD helicase. *Mol. Cell,* **35**, 694–703.

[169] Rueda, D. and Walter, N. G. (2005) Single molecule fluorescence control for nanotechnology. *J. Nanosci. Nanotechnol.,* **5**, 1–11.

[170] Zhao, R. and Rueda, D. (2009) RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods,* **49**, 112–117.

[171] Edidin, M. (2001) Near-field scanning optical microscopy, a siren call to biology. *Traffic,* **2**, 797–803.

[172] Tokunaga, M., Inmamoto, N., and Sakata-Sogawa, K. (2008) Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods,* **5**, 159–161.

[173] Wolf, K. B. and Krötzsch, G. (1995) Geometry and dynamics in refracting systems. *Eur. J. Phys.,* **16**, 14–20.

[174] Betzig, E. and Chichester, R. J. (1993) Single molecules observed by near-field scanning optical microscopy. *Science,* **262**, 1422–1425.

[175] Ha, T., Enderle, T., F, O. D., Chemla, D. S., Selvin, P. R., and Weiss, S. (1996) Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. USA,* **93**, 6264–6268.

[176] Ha, T., Zhuang, X., Kim, H. D., Orr, J. W., Williamson, J. R., and Chu, S. (1999) Ligand-induced conformational changes observed in single RNA molecules. *Proc. Natl. Acad. Sci. USA,* **96**, 9077–9082.

[177] Dunn, R. C. (1999) Near-field scanning optical microscopy. *Chem. Rev.,* **99**, 2891–2927.

[178] Ritter, J. G., Veith, R., Veenendaal, A., Siebrasse, J. P., and Kubitscheck, U. (2010) Light sheet microscopy for single molecule tracking in living tissue. *PLoS ONE,* **5**, e11639.

[179] Coelho, M., Maghelli, N., and Tolic-Norrelykke, I. M. (2013) Single-molecule imaging *in vivo*: the dancing building blocks of the cell. *Integr. Biol.,* **5**, 748–758.

[180] Haustein, E. and Schwille, P. (2007) Fluorescence correlation spectroscopy: novel variations of an established technique. *Annu. Rev. Biophys. Biomol. Struct.,* **36**, 151–169.

[181] Forkey, J. N., Quinlan, M. E., and Goldman, Y. E. (2000) Protein structural dynamics by single-molecule fluorescence polarization. *Prog. Biophys. Mol. Biol.,* **74**, 1–35.

[182] Bastiaens, P. I. H. and Squire, A. (1999) Fluorescence lifetime imaging microscopy: spatial resolution of biochemical processes in the cell. *Trends Cell Biol.,* **9**, 48–52.

[183] Deniz, A. A., Dahan, M., Grunwell, J. R., Ha, T., Faulhaber, A. E., Chemla, D. S., Weiss, S., and Schultz, P. G. (1999) Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Förster distance dependence and subpopulations. *Proc. Natl. Acad. Sci. USA,* **96**, 3670–3675.

[184] Hirschfeld, V., Paulsen, H., and Hübner, C. G. (2013) The spectroscopic ruler revisited at 77 K. *Phys. Chem. Chem. Phys.,* **15**, 17664–17671.

[185] Hoffmann, A., Nettels, D., Clark, J., Borgia, A., Radford, S. E., Clarke, J., and Schuler, B. (2011) Quantifying heterogeneity and conformational dynamics from single molecule FRET of diffusing molecules: recurrence analysis of single particles (RASP). *Phys. Chem. Chem. Phys.,* **13**, 1857–1871.

[186] Picoquant GmbH. (Oct 2013) Photon counting detectors, http://www.picoquant.com/products/.

[187] Beard, S. M. and MacGillivray, H. T. (1990) The COSMOS system for crowded-field analysis of digitized photographic plate scans. *Mon. Not. R. astr. Soc.,* **247**, 311–321.

[188] Duda, R. O. and Hart, P. E. (1972) Use of the Hough transformation to detect lines and curves in pictures. *Comm. ACM,* **15**, 11–15.

[189] Holden, S. J., Uphoff, S., Hohlbein, J., Yadin, D., Le Reste, L., Britton, O. J., and Kapanidis, A. N. (2010) Defining the limits of single-molecule FRET resolution in TIRF microscopy. *Biophys. J.,* **99**, 3102–3111.

[190] Rebull, L. M., Johnson, C. H., Hoette, V., Kim, J. S., Laine, S., Foster, M., Laher, R., Legassie, M., Mallory, C. R., McCarron, K., and Sherry, W. H. (2011) New Young Star Candidates in CG4 and Sa101. *Astron. J.,* **142**, 25–64.

[191] Lee, N. K., Kapanidis, A. N., Wang, Y., Michalet, X., Mukhopadhyay, J., Ebright, R. H., and Weiss, S. (2005) Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.,* **88**, 2939–2953.

[192] McKinney, S. A., Joo, C., and Ha, T. J. (2006) Analysis of single molecule FRET trajectories using hidden Markov modeling. *Biophys. J.,* **91**, 1941–1951.

[193] Helm, M., Kobitski, A. Y., and Nienhaus, G. U. (2009) Single-molecule Förster resonance energy transfer studies of RNA structure, dynamics and function. *Biophys. Rev.,* **1**, 161–176.

[194] Pljevaljcic, G., Millar, D. P., and Deniz, A. A. (2004) Freely diffusing single hairpin ribozymes provide insights into the role of secondary structure and partially folded states in RNA folding. *Biophys. J.,* **87**, 457–467.

[195] Zhuang, X., Kim, H., Pereira, M. J. B., Babcock, H. P., Walter, N. G., and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science,* **296**, 1473–1476.

[196] Bokinsky, G., Rueda, D., Misra, V. K., Rhodes, M. M., Gordus, A., Babcock, H. P., Walter, N. G., and Zhuang, X. (2003) Single-molecule transition-state analysis of RNA folding. *Proc. Natl. Acad. Sci. USA,* **100**, 9302–9307.

[197] Rueda, D., Bokinsky, G., Rhodes, M. M., Rust, M. J., Zhuang, X., and Walter, N. G. (2004) Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proc. Natl. Acad. Sci. USA,* **101**, 10066–1071.

[198] Ditzler, M. A., Rueda, D., Mo, J., Hakansson, K., and Walter, N. G. (2008) A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res.,* **36**, 7088–7099.

[199] Tan, E., Wilson, T. J., Nahas, M. K., Clegg, R. M., Lilley, D. M. J., and Ha, T. (2003) A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proc. Natl. Acad. Sci. USA,* **100**, 9308–9313.

[200] Okumus, B., Wilson, T. J., Lilley, D. M. J., and Ha, T. (2004) Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys. J.,* **87**, 2798–2806.

[201] Hohng, S., Wilson, T. J., Tan, E., Clegg, R. M., Lilley, D. M., and Ha, T. (2004) Conformational flexibility of four-way junctions in RNA. *J. Mol. Biol.,* **336**, 69–79.

[202] Nahas, M. K., Wilson, T. J., Hohng, S., Jarvie, K., Lilley, D. M., and Ha, T. (2004) Observation of internal cleavage and ligation reactions of a ribozyme. *Nat. Struct. Mol. Biol.,* **11**, 1107–1113.

[203] Liu, S., Bokinsky, G., Walter, N. G., and Zhuang, X. (2007) Dissecting the multistep reaction pathway of an RNA enzyme by single-molecule kinetic fingerprinting. *Proc. Natl. Acad. Sci. USA,* **104**, 12634–12639.

[204] Steiner, M., Karunatilaka, K. S., Sigel, R. K. O., and Rueda, D. (2008) Single molecule studies of group II intron ribozymes. *Proc. Natl. Acad. Sci. USA,* **105**, 13853–13858.

[205] Karunatilaka, K. S., Solem, A., Pyle, A. M., and Rueda, D. (2010) Single-molecule analysis of Mss116-mediated group II intron folding. *Nature,* **467**, 935–939.

[206] Stone, M. D., Mihalusova, N., O'Connor, C. M., Prathapam, R., Collins, K., and Zhuang, X. (2007) Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature,* **446**, 458–461.

[207] Pereira, M. J. B., Nikolova, E. N., Hiley, S. L., Jaikaran, D., Collins, R. A., and Walter, N. G. (2008) Single VS ribozyme molecules reveal dynamic and hierarchical folding toward catalysis. *J. Mol. Biol.,* **382**, 496–509.

[208] Wozniak, A. K., Nottrott, S., Kühn-Hölsken, E., Schröder, G. F., Grubmüller, H., Lührmann, R., Seidel, C. A. M., and Oesterheld, F. (2005) Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements. *RNA,* **11**, 1545–1554.

[209] Guo, Z., Karunatilaka, K. S., and Rueda, D. (2009) Single-molecule analysis of protein-free U2-U6 snRNAs. *Nat. Struct. Mol. Biol.,* **16**, 1154–1160.

[210] Karunatilaka, K. S. and Rueda, D. (2014) Post-transcriptional modifications modulate conformational dynamics in human U2-U6 snRNA complex. *RNA,* **20**, 16–23.

[211] Lemay, J. F., Penedo, J. C., Tremblay, R., and Lilley, D. M. (2006) Folding of the adenine riboswitch. *Chem. Biol.,* **13**, 857–868.

[212] Fiegland, L. R., Garst, A. D., Batey, R. T., and Nesbitt, D. J. (2012) Single-molecule studies of the lysine riboswitch reveal effector dependent conformational dynamics of the aptamer domain single-molecule dependent conformational dynamics of the aptamer domain. *Biochemistry,* **51**, 9223–9233.

[213] Wood, S., Ferré-D'Amaré, A., and Rueda, D. (2012) Allosteric interactions preorganize the c-di-GMP riboswitch and accelerate ligand binding. *ACS Chem. Biol.,* **7**, 907–914.

[214] Zhao, R., Marshall, M., Aleman, E. A., Lamichhane, R., Feig, A., and Rueda, D. (2010) Laser-assisted single-Molecule refolding (LASR). *Biophys. J.,* **99**, 1925–1931.

[215] Salim, N., Lamichhane, R., Zhao, R., Banerjee, T., Philip, J., Rueda, D., and Feig, A. L. (2012) Thermodynamic and kinetic analysis of an RNA kissing interaction and its resolution into an extended hairpin. *Biophys. J.,* **102**, 1097–1107.

[216] Voigts-Hoffmann, F., Hengesbach, M., Kobitski, M. Y., van Aerschot, A., Herdewijn, P., Nienhaus, G. U., and Helm, M. (2007) A methyl group controls conformational equilibrium in human mitochondrial tRNA$^{Lys}$. *J. Am. Chem. Soc.,* **129**, 13382–13383.

[217] Kobitski, A. Y., Hengesbach, M., Helm, M., and Nienhaus, U. G. (2008) Sculpting an RNA conformational energy landscape by a methyl group modification - a single-molecule FRET study. *Angew. Chem. Int. Ed.,* **47**, 4326–4330.

[218] Daher, M. and Rueda, D. (2012) Fluorescence of the transfer RNA-like domain of transfer messenger RNA in complex with small binding protein B. *Biochemistry,* **51**, 3531–3538.

[219] Zhuang, X., Bartley, L. E., Babcock, H. P., Russel, R., Ha, T., Herschlag, D., and Chu, S. (2000) A single-molecule study of RNA catalysis and folding. *Science,* **288**, 2048–2051.

[220] Russell, R., Zhuang, X., Babcock, H. P., Millet, I. S., Doniach, S., Chu, S., and Herschlag, D. (2002) Exploring the folding landscape of a structured RNA. *Proc. Natl. Acad. Sci. USA,* **99**, 155–160.

[221] Bartley, L. E., Zhuang, X., Das, R., Chu, S., and Herschlag, D. (2003) Exploration of the transition state for tertiary structure formation between an RNA helix and a large structured RNA. *J. Mol. Biol.,* **328**, 1011–1026.

[222] Lee, T. H., Lapidus, L. J., Zhao, W., Travers, K. J., Herschlag, D., and Chu, S. (2007) Measuring the folding transition time of single RNA molecules. *Biophys. J.,* **92**, 3275–3283.

[223] Sattin, B. D., Zhao, W., Travers, K., Chu, S., and Herschlag, D. (2008) Direct measurement of tertiary contact cooperativity in RNA folding. *J. Am. Chem. Soc.,* **130**, 6085–6087.

[224] Solomatin, S. V., Greenfield, M., and Herschlag, D. (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature,* **463**, 681–684.

[225] Solomatin, S. V., Greenfield, M., and Herschlag, D. (2011) Implications of molecular heterogeneity for the cooperativity of biological macromolecules. *Nat. Struct. Mol. Biol.,* **18**, 732–734.

[226] Greenfield, M., Solomatin, S. V., and Herschlag, D. (2011) Removal of covalent heterogeneity reveal simple folding behavior for P4-P6. *J. Biol. Chem.,* **286**, 19872–19879.

[227] Bokinsky, G., Nivon, L. G., Liu, S., Chai, G., Hong, M., Weeks, K. M., and Zhuang, X. (2006) Two distinct binding modes of a protein cofactor with its target RNA. *J. Mol. Biol.,* **361**, 771–784.

[228] Xie, S., Srividya, N., Sosnick, T. R., Pan, T., and Scherer, N. F. (2004) Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme. *Proc. Natl. Acad. Sci. USA,* **101**, 534–539.

[229] Smith, G. J., Lee, K. T., Qu, X., Xie, Z., Pesic, J., Sosnick, T. R., and Scherer, N. F. (2008) A large collapsed-state RNA can exhibit simple exponential single-molecule dynamics. *J. Mol. Biol.,* **378**, 943–953.

[230] Qu, X., Smith, G. J., Lee, K. T., Sosnick, T. R., Pan, T., and Scherer, N. F. (2008) Single-molecule nonequilibrium periodic $Mg^{2+}$-concentration jump experiments reveal details of the early folding pathways of a large RNA. *Proc. Natl. Acad. Sci. USA,* **105**, 6602–6607.

[231] Elenko, M. P., Szostak, J. W., and van Oijen, A. M. (2009) Single-molecule imaging of an *in vitro*-evolved RNA aptamer reveals homogeneous ligand binding kinetics. *J. Am. Chem. Soc.,* **131**, 9866–9867.

[232] De Silva, C. and Walter, N. G. (2009) Leakage and slow allostery limit performance of single drug-sensing aptazyme molecules based of the hammerhead ribozyme. *RNA,* **15**, 76–84.

[233] Kobitski, A. Y., Nierth, A., Helm, M., Jaschke, A., and Nienhaus, G. U. (2007) $Mg^{2+}$-dependent folding of a Diels-Alderase ribozyme probed by single-molecule FRET analysis. *Nucleic Acids Res.,* **35**, 2047–2059.

[234] Doetsch, M., Schroeder, R., and Fürtig, B. (2011) Transient RNA-protein interactions in RNA folding. *FEBS J.,* **278**, 1634–1642.

[235] Garneau, N. L., Wilusz, J., and Wilusz, C. J. (2007) The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell. Biol.,* **8**, 113–126.

[236] Ibba, M. and Söll, D. (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.,* **69**, 617–650.

[237] Proudfoot, N. J., Furger, A., and Dye, M. J. (2002) Integrating mRNA processing with transcription. *Cell,* **108**, 501–512.

[238] Yusupov, M. M., Yusupova, G. Z., Baucom, A., Liebermann, K., Earnerst, T. N., Cate, J. H. D., and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science,* **292**, 883–896.

[239] Hoskins, A. A. and Moore, M. J. (2012) The spliceosome: a flexible, reversible macromolecular structure. *Trends Biochem. Sci.,* **37**, 179–188.

[240] van der Feltz, C., Anthony, K., Brilot, A., and Pomeranz Krummel, D. A. (2012) Architecture of the spliceosome. *Biochemistry,* **51**, 3321–3333.

[241] Williams, G. T. and Farzaneh, F. (2012) Are snoRNAs and SnoRNA host genes new players in cancer. *Nat. Rev.,* **12**, 84–88.

[242] Feng, J., Funk, W. D., Wang, S. S., Weindrich, S. L., Avilion, A. A., Chiu, C. P., Adams, R. R., Chang, E., Allsopp, R. C., Yu, J., Le, S., West, M. D., Harley, C. B., Andrews, W. H., Greider, C. W., and Villeponteau, B. (1995) The RNA component of human telomerase. *Science,* **269**, 1236–1241.

[243] Brantl, S. (2007) Regulatory mechanisms employed by *cis*-encoded antisense RNAs. *Curr. Opin. Microbiol.,* **10**, 102–109.

[244] Bartel, D. P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell,* **136**, 215–233.

[245] Ferré-D'Amaré, A. and Winkler, W. C. (2011) The roles of metal ions in regulation by riboswitches. *Met. Ions Life Sci.,* **9**, 141–173.

[246] Hamilton, A. J. and Baucombe, D. C. (1999) A species of small antisense RNA in post-transcriptional gene silencing in plants. *Science,* **286**, 950–952.

[247] Montange, R. K. and Batey, R. T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.,* **37**, 117–133.

[248] Krupovic, A. and Cvirkaite-Krupovic, V. (2011) Virophages or satellite viruses. *Nat. Rev. Microbiol.,* **9**, 762–763.

[249] Simon-Loriere, E. and Holmes, E. C. (2011) Why do RNA viruses recombine. *Nat. Rev. Microbiol.,* **9**, 617–626.

[250] Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMigue, l. P., and Schulman, A. H. (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.,* **8**, 973–982.

[251] Butcher, S. E. and Pyle, A. M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.,* **44**, 1302–1311.

[252] Huppert, J. L., Bugaut, A., Kumari, S., and Balasubramanian, S. (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.,* **36**, 6260–6268.

[253] Pyle, A. M. (2002) Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem.,* **7**, 679–690.

[254] Treiber, D. K. and Williamson, J. R. (2001) Beyond kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.,* **11**, 309–314.

[255] Williamson, J. R. (2008) Cooperativity in macromolecular assembly. *Nat. Chem. Biol.,* **4**, 458–465.

[256] Woodson, S. A. (2005) Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.,* **9**, 104–109.

[257] Herschlag, D. (1995) RNA chaperones and the RNA folding problem. *J. Biol. Chem.,* **270**, 20871–20874.

[258] Pyle, A. M. (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.,* **37**, 317–336.

[259] Gingras, A. C., Raught, B., and Sonenberg, N. (2001) Regulation of translation initiation by FRAP/mTOR. *Gene Dev.,* **15**, 807–826.

[260] Parsyan, A., Shahbazian, D., Martineau, Y., Petroulakis, E., Alain, T., Larsson, O., Mathonnet, G., Tettweiler, G., Hellen, C. U., Pestova, T. V., Svitkin, Y. V., and Sonenberg, N. (2009) The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis. *Proc. Natl. Acad. Sci. USA,* **106**, 22217–22222.

[261] Fromont-Racine, M., Senger, B., Saveanu, C., and Fasiolo, F. (2003) Ribosome assembly in eukaryotes. *Gene,* **313**, 17–42.

[262] Mayer, C. and Grummt, I. (2006) Ribosome biogenesis and cell growth: mTOR coordinates transcription by all three classes of nuclear RNA polymerases. *Oncogene,* **25**, 6384–6391.

[263] Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S., and Wu, J. Y. (2011) RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5’ splice site. *Mol. Cell. Biol.,* **31**, 1812–1821.

[264] Liu, Z. R. (2002) p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex. *Mol. Cell. Biol.,* **22**, 5443–5450.

[265] Linder, P. and Jankowsky, E. (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell. Biol.,* **12**, 505–516.

[266] Newman, M. A. and Hammond, S. M. (2010) Emerging paradigms of regulated microRNA processing. *Gene Dev.,* **24**, 1086–1092.

[267] Winter, J., Jung, S., Keller, S., and Gregory, R. I. (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell. Biol.,* **11**, 228–234.

[268] Gong, Z., Dong, C. H., Lee, H., Zhu, J., Xiong, L., Gong, D., Stevenson, B., and Zhu, J. K. (2005) A DEAD box RNA helicase is essential for mRNA export and important for development and stress responses in *Arabidopsis*. *Plant Cell,* **17**, 256–267.

[269] Lai, M. C., Wu Lee, Y. H., and Tarn, W. Y. (2008) The DEAD-box RNA helicase DDX3 associates with export messenger ribonucleoproteins as well as tip-associated protein and participates in translational control. *Mol. Biol. Cell.,* **19**, 3847–3858.

[270] Ranji, A. and Boris-Lawrie, K. (2010) RNA helicases: emerging roles in viral replication and the host innate response. *RNA Biol.,* **7**, 775–787.

[271] Lattmann, S., Stadler, M. B., Vaughn, J. P., Akman, S. A., and Nagamine, Y. (2011) The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. *Nucleic Acids Res.,* **39**, 9390–9404.

[272] Henn, A., Bradley, and De la Cruz, E. M. (2012) ATP utilization and RNA conformational rearrangement by DEAD-box proteins. *Annu. Rev. Biophys.,* **41**, 247–267.

[273] Cao, W., Coman, M. M., Ding, S., Henn, A., Middleton, E. R., Bradley, M. J., Rhoades, E., Hackney, D. D., Pyle, A. M., and De la Cruz, E. M. (2011) Mechanism of Mss116 ATPase reveals functional diversity of DEAD-box proteins. *J. Mol. Biol.,* **409**, 399–414.

[274] Henn, A., Cao, W., Licciardello, N., Heitkamp, S. E., Hackney, D. D., and De la Cruz, E. M. (2010) Pathway of ATP utilization and duplex rRNA unwinding by the DEAD-box helicase DbpA. *Proc. Natl. Acad. Sci. USA,* **107**, 4046–4050.

[275] Singleton, M. R., Dillingham, M. S., and Wigley, D. B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.,* **76**, 23–50.

[276] Jankowsky, E. and Fairman, M. E. (2007) RNA helicases - one fold for many functions. *Curr. Opin. Chem. Biol.,* **17**, 316–324.

[277] Pyle, A. M. (2011) RNA helicases and remodeling proteins. *Curr. Opin. Chem. Biol.,* **15**, 636–642.

[278] Andreou, A. Z. and Klostermeier, D. (2012) Conformational changes of DEAD-box helicases monitored by single molecule fluorescence resonance energy transfer. *Methods Enzymol.,* **511**, 75–109.

[279] Tai, C. L., Chi, W. K., Chen, D. S., and Hwang, L. H. (1996) The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). *J. Virol.,* **70**, 8477–8484.

[280] Eggleston, A. K., Rahim, N. A., and Kowalczykowski, S. C. (1996) A helicase assay based on the displacement of fluorescent, nucleic acid-binding ligands. *Nucleic Acids Res.,* **24**, 1179–1186.

[281] Tseng-Rogenski, S. S. and Chang, T. H. (2004) RNA unwinding assay for DExD/H-box RNA helicases. *Methods Enzymol.,* **257**, 93–102.

[282] Yodh, J. G., Schlierf, M., and Ha, T. (2010) Insights into helicase mechanism and function revealed through single-molecule approaches. *Q. Rev. Biophys.,* **43**, 185–217.

[283] Gordon, T., Ha, T., and Selvin, P. R. (2004) Single-molecule high-resolution imaging with photobleaching. *Proc. Natl. Acad. Sci. USA,* **101**, 6462–6465.

[284] Blanco, M. and Walter, N. G. (2010) Analysis of complex single molecule FRET time trajectories. *Methods Enzymol.,* **472**, 153–178.

[285] Diao, J., Su, Z., Lu, X., Yoon, T. Y., Shin, Y. K., and Ha, T. (2010) Single-vesicle fusion assay reveals Munc18-1 binding to the SNARE core is sufficient for stimulating membrane fusion. *ACS Chem. Neuro.,* **1**, 168–174.

[286] Uphoff, S., Holden, S. J., Le Reste, L., Periz, J., van de Linde, S., Heilemann, M., and Kapanidis, A. N. (2010) Monitoring multiple distances within a single molecule using switchable FRET. *Nat. Methods,* **7**, 831–836.

[287] Chung, H. S., Gopich, I. V., McHale, K., Cellmer, T., Louis, J. M., and Eaton, W. A. (2011) Extracting rate coefficients from single-molecule photon trajectories and FRET efficiency histograms for a fast-folding protein. *J. Phys. Chem. A,* **115**, 3642–3656.

[288] Neuman, K. C. and Nagy, A. (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods,* **5**, 491–505.

[289] Karunatilaka, K. S. and Rueda, D. (2009) Single-molecule fluorescence studies of RNA: a decade's progress. *Chem. Phys. Lett.,* **476**, 1–10.

[290] Cardo, L., Karunatilaka, K. S., Rueda, D., and Sigel, R. K. O. Single molecule FRET characterization of large ribozyme folding. *In*: Ribozymes: methods and protocols pp. 227–251 Humana Press Inc., Totowa, NJ, USA (2012).

[291] Andreou, A. Z. and Klostermeier, D. (2013) The DEAD-box helicase eIF4A: paradigm or the odd one out. *RNA Biol.,* **10**, 19–32.

[292] Hopfner, K. P. and Michaelis, J. (2007) Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics. *Curr. Opin. Struct. Biol.,* **17**, 87–95.

[293] Rasnik, I., Myong, S., and Ha, T. (2006) Unraveling helicase mechanisms one molecule at a time. *Nucleic Acids Res.,* **34**, 4225–4231.

[294] Steimer, L. and Klostermeier, D. (2012) RNA helicases in infection and disease. *RNA Biol.,* **9**, 751–771.

[295] Tinoco, I., Chen, G., and Qu, X. RNA worlds: from life's origins to diversity in gene regulation. *In*: Cold spring harbor perspectives in biology pp. 227–251 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA (2011).

[296] Dholakia, K., Reece, P., and Gu, M. (2008) Optical micromanipulation. *Chem. Soc. Rev.,* **37**, 42–55.

[297] Katan, A. J. and Dekker, C. (2011) High-speed AFM reveals the dynamics of single biomolecules at the nanometer scale. *Cell,* **147**, 979–982.

[298] Campos, L. A., Liu, J., Wang, X., Ramanathan, R., English, D. S., and Munoz, V. (2011) A photoprotection strategy for microsecond-resolution single-molecule fluorescence spectroscopy. *Nat. Methods,* **8**, 143–146.

[299] Hwang, H., Kim, H., and Myong, S. (2011) Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. *Proc. Natl. Acad. Sci.,* **108**, 7414–7418.

[300] Greenleaf, W. J., Woodside, M. T., and Block, S. M. (2007) High-resolution, single-molecule measurements of biomolecular motion. *Annu. Rev. Biophys. Biomol. Struct.,* **36**, 171–190.

[301] Henn, A., Medalia, O., Shi, S. P., Steinberg, M., Franceschi, F., and Sagi, I. (2001) Visualization of unwinding activity of duplex RNA by DbpA, a DEAD box helicase, at single-molecule resolution by atomic force microscopy. *Proc. Natl. Acad. Sci. USA,* **98**, 5007–2012.

[302] Marsden, S., Nardelli, M., Linder, P., and McCarthy, J. E. G. (2006) Unwinding single RNA molecules using helicases in eukaryotic translation initiation. *J. Mol. Biol.,* **361**, 327–335.

[303] Dohoney, K. M. and Gelles, J. (2000) X-sequence recognition and DNA translocation by single RecBCD helicase/nuclease molecules. *Nature,* **409**, 370–374.

[304] Fan, H. F. and Li, H. W. (2009) Studying RecBCD helicase translocation along X-DNA using tethered particle motion with a stretching force. *Biophys. J.,* **96**, 1875–1883.

[305] Dessinges, M. N., Lionnet, T., Xi, X. G., Bensimon, D., and Croquette, V. (2004) Single-molecule assay reveals strand switching and enhanced processivity of UvrD. *Proc. Natl. Acad. Sci. USA,* **101**, 6439–6444.

[306] Sun, B., Wei, K. J., Zhang, B., Zhang, X. H., Dou, S. X., Li, M., and Xi, X. G. (2008) Impediment of *E. coli* UvrD by DNA-destabilizing force reveals a strained-inchworm mechanism of DNA unwinding. *EMBO J.,* **27**, 3279–3287.

[307] Binnig, G., Quate, C. F., and Gerber, C. (1986) Atomic force microscopy. *Phys. Rev. Lett.,* **56**, 930–933.

[308] Cohen, S. R. and Bitler, A. (2008) Use of AFM in bio-related systems. *Curr. Opin. Colloid. Interface Sci.,* **13**, 316–325.

[309] Heus, H. A., Puchner, E. M., van Vugt-Jonker, A. J., Zimmermann, J. L., and Gaub, H. E. (2011) Atomic force microscope-based single-molecule force spectroscopy of RNA unfolding. *Anal. Biochem.,* **414**, 1–6.

[310] Santos, N. C. and Castanho, M. A. R. B. (2004) An overview of the biophysical applications of atomic force microscopy. *Biophys. Chem.,* **107**, 133–149.

[311] Vilfan, I. D., Kamping, W., van den Hout, M., Candelli, A., Hage, S., and Dekker, N. H. (2007) An RNA toolbox for single-molecule force spectroscopy studies. *Nucleic Acids Res.,* **35**, 6625–6639.

[312] Neaves, K. J., Huppert, J. L., Henderson, R. M., and Edwardson, J. M. (2009) Direct visualization of G-quadruplexes in DNA using atomic force microscopy. *Nucleic Acids Res.,* **37**, 6269–6275.

[313] Picco, L. M., Bozec, L., Ulcinas, A., Engledew, D. J., Antognozzi, M., Horton, M. A., and Miles, M. J. (2007) Breaking the speed limit with atomic force microscopy. *Nanotechnology,* **18**, 044030.

[314] Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E., and Chu, S. (1986) Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.,* **11**, 288–290.

[315] Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R., and M, B. S. (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature,* **438**, 460–465.

[316] Herbert, K. M., La Porta, A., Wong, B. J., Mooney, R. A., Neuman, K. C., Landick, R., and Block, S. M. (2006) Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell,* **125**, 1083–1094.

[317] Rohrbach, A. (2005) Switching and measuring a force of 25 femtoNewtons with an optical trap. *Opt. Express,* **13**, 9695–9701.

[318] Shaevitz, J. W., Abbondanzieri, E. A., Landick, R., and Block, S. M. (2003) Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature,* **426**, 684–687.

[319] Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009) Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science,* **325**, 626–628.

[320] Wang, M. D., Yin, H., Landick, R., Gelles, J., and Block, S. M. (1997) Stretching DNA with optical tweezers. *Biophys. J.,* **72**, 1335–1346.

[321] Finer, J. T., Simmons, R. M., and Spudich, J. A. (1994) Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature,* **368**, 113–119.

[322] Svoboda, K., Schmidt, C. F., Schnapp, B. J., and Block, S. M. (1993) Direct observation of kinesin stepping by optical trapping interferometry. *Nature,* **365**, 721–727.

[323] Hendricks, A. G., Holzbaur, E. L. F., and Goldman, Y. E. (2012) Force measurements on cargoes in living cells reveal collective dynamics of microtubule motors. *Proc. Natl. Acad. Sci. USA,* **109**, 18447–18452.

[324] Block, S. M. (1992) Making light work with optical tweezers. *Nature,* **360**, 493–495.

[325] Kozak, M. (1986) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA,* **83**, 2850–2854.

[326] Kozak, M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene,* **361**, 13–37.

[327] Vega Laso, M. R., Zhu, D., Sagliocco, F., Brown, A. J. P., Tuite, M. F., and McCarthy, J. E. G. (1993) Inhibition of translational initiation in the yeast *Saccharomyces cerevisiae* as a function of the stability and position of hairpin structures in the mRNA leader. *J. Biol. Chem.,* **268**, 6453–6462.

[328] Rogers Jr., G. W., Richter, N. J., and Merrick, W. C. (1999) Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J. Biol. Chem.,* **274**, 12236–12244.

[329] Iost, I., Dreyfus, M., and Linder, P. (1999) Ded1p, a DEAD-box protein required for translation initiation in *Saccharomyces cerevisiae*, is an RNA helicase. *J. Biol. Chem.,* **274**, 17677–17683.

[330] Hilbert, M., Kebbel, F., Gubaev, A., and Klostermeier, D. (2011) eIF4G stimulates the activity of the DEAD box protein eIF4A by a conformational guidance mechanism. *Nucleic Acids Res.,* **39**, 2260–2270.

[331] Dimitrova, M., Imbert, I., Kieny, M. P., and Schuster, C. (2003) Protein-protein interactions between hepatitis C virus nonstructural proteins. *J. Virol.,* **77**, 5401–5414.

[332] Pang, P. S., Jankowsky, E., Planet, P. J., and Pyle, A. M. (2002) The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. *EMBO J.,* **21**, 1168–1176.

[333] Dumont, S., Cheng, W., Serebrov, V., Beran, R. K., Tinoco, I., Pyle, A. M., and Bustamante, C. (2006) RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. *Nature,* **439**, 105–108.

[334] Cheng, W., Dumont, S., Tinoco, I., and Bustamante, C. (2007) NS3 helicase actively separates RNA strands and senses sequence barriers ahead of the opening fork. *Proc. Natl. Acad. Sci. USA,* **104**, 13954–13959.

[335] Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. (1994) Entropic elasticity of lambda-phage DNA. *Science,* **265**, 1599–1600.

[336] Myong, S., Bruno, M. M., Pyle, A. M., and Ha, T. (2007) Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. *Science,* **317**, 513–516.

[337] Cheng, W., Arunajadai, S. G., Moffit, J. R., Tinoco, I. J., and Bustamante, C. (2011) Single-base pair unwinding and asynchronous RNA release by the hepatitis C virus NS3 helicase. *Science,* **333**, 1746–1749.

[338] Serebrov, V. and Pyle, A. M. (2004) Periodic cycles of RNA unwinding and pausing by hepatitis C virus NS3 helicase. *Nature,* **430**, 476–480.

[339] Appleby, T. C., Anderson, R., Fedorova, O., Pyle, A. M., Wang, R., Liu, X., Brendza, K. M., and R, S. J. (2011) Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV. *J. Mol. Biol.,* **405**, 1139–1153.

[340] Fairman-Williams, M. E. and Jankowsky, E. (2012) Unwinding initiation by the viral RNA helicase NPH-II. *J. Mol. Biol.,* **415**, 819–832.

[341] Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P. J., Berger, S., Restle, T., Goody, R. S., Gohlke, H., and Seidel, C. A. M. (2012) A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods,* **9**, 1218–1225.

[342] Orte, A., Clarke, R., and Klenerman, D. (2010) Single-molecule two-colour coincidence detection to probe biomolecular associations. *Biochem. Soc. Trans.,* **38**, 914–918.

[343] Yang, Q. and Jankowsky, E. (2006) The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat. Struct. Biol.,* **13**, 981–986.

[344] Aregger, R. and Klostermeier, D. (2009) The DEAD box helicase YxiN maintains a closed conformation during ATP hydrolysis. *Biochemistry,* **48**, 10679–10681.

[345] Karow, A. R. and Klostermeier, D. (2009) A conformational change in the helicase core is necessary but not sufficient for RNA unwinding by the DEAD box helicase YxiN. *Nucleic Acids Res.,* **37**, 4464–4471.

[346] Linden, M. H., Hartmann, R. K., and Klostermeier, D. (2008) The putative RNase P motif in the DEAD box helicase Hera is dispensable for efficient interaction with RNA and helicase activity. *Nucleic Acids Res.,* **36**, 5800–5811.

[347] Theissen, B., Karow, A. R., Köhler, J., Gubaev, A., and Klostermeier, D. (2008) Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase. *Proc. Natl. Acad. Sci. USA,* **105**, 548–553.

[348] Vogelsang, J., Kasper, R., Steinhauer, C., Person, B Heilemann, M., Sauer, M., and Tinnefeld, P. (2008) A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chem. Int. Ed.,* **47**, 5465–5469.

[349] Myong, S., Cui, S., Cornish, P. V., Kirchhofer, A., Gack, M. U., Jung, J. U., Hopfner, K. P., and Ha, T. (2009) Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. *Science,* **323**, 1070–1074.

[350] Markiewicz, R. P., Vritis, K. B., Rueda, D., and Romano, L. J. (2012) Single-molecule microscopy reveals new insights into nucleotide selection by DNA polymerase I. *Nucleic Acids Res.,* **40**, 7975–7984.

[351] Staley, J. P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell,* **92**, 315–326.

[352] Schroeder, R., Barta, A., and Semrad, K. (2004) Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell. Biol.,* **5**, 908–919.

[353] Yang, Q., Fairman, M. E., and Jankowsky, E. (2007) DEAD-box-protein-assisted RNA structure conversion towards and against thermodynamic equilibrium values. *J. Mol. Biol.,* **368**, 1087–1100.

[354] Kapanidis, A. N., Lee, N. K., Laurence, T. A., Margeat, E., and Weiss, S. (2004) Fluorescence-aided molecules sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. USA,* **101**, 8936–8941.

[355] Kapanidis, A. N., Laurence, T. A., Lee, N. K., Margeat, E., Kong, X., and Weiss, S. (2005) Alternating-laser excitation of single molecules. *Acc. Chem. Res.,* **38**, 523–533.

[356] Kim, D. W., Gwack, Y., Han, J. H., and Choe, J. (1995) C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. *Biochem. Biophys. Res. Commun.,* **215**, 160–166.

[357] Gross, C. H. and Shuman, S. (1996) *Vaccinia* virions lacking the RNA helicase nucleoside triphosphate phosphohydrolase II are defective in early transcription. *J. Virol.,* **70**, 8549–8557.

[358] Jankowsky, E., Gross, C. H., Shuman, S., and Pyle, A. M. (2000) The DExH protein NPH-II is a processive and directional motor for unwinding RNA. *Nature,* **403**, 447–451.

[359] Shuman, S. (1993) *Vaccinia* virus RNA helicase. Directionality and substrate specificity. *J. Biol. Chem.,* **268**, 11798–11802.

[360] Wong, I. and Lohman, T. M. (1997) A two-site mechanism for ATP hydrolysis by the asymmetric Rep dimer P2S as revealed by site-specific inhibition with ADP-AlF4. *Biochemistry,* **36**, 3115–3125.

[361] Levin, M. K., Gurjar, M. M., and Patel, S. S. (2003) ATP binding modulates the nucleic acid affinity of hepatitis C virus helicase. *J. Biol. Chem.,* **278**, 23311–23316.

[362] Chen, B., Doucleff, M., Wemmer, D. E., De Carlo, S., Huang, H. H., Nogales, E., Hoover, T. R., Kondrashkina, E., Guo, L., and Nixon, B. T. (2007) ATP group- and transition-states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, sigma54. *Structure,* **15**, 429–440.

[363] Liu, F., Putnam, A., and Jankowsky, E. (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci. USA,* **105**, 20209–20214.

[364] Gu, M. and Rice, C. M. (2010) Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl. Acad. Sci. USA,* **107**, 521–528.

[365] Chakrabarti, S., Jaychandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., and Conti, E. (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell,* **41**, 693–703.

[366] Kawaoka, J., Jankowsky, E., and Pyle, A. M. (2004) Backbone tracking by the SF2 helicase NPH-II. *Nat. Struct. Mol. Biol.,* **11**, 526–530.

[367] Kawaoka, J. and Pyle, A. M. (2005) Choosing between DNA and RNA: the polymer specificity of RNA helicase NPH-II. *Nucleic Acids Res.,* **33**, 644–649.

[368] Halls, C., Mohr, S., Del Campo, M., Yang, Q., Jankowsky, E., and Lambowitz, A. M. (2007) Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. *J. Mol. Biol.,* **365**, 835–855.

[369] Ha, T., Rasnik, I., Cheng, W., Babcock, H. P., Gauss, G. H., Lohman, T. M., and Chu, S. (2002) Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature,* **419**, 638–641.

[370] Lamichhane, R., Solem, A., Black, W., and Rueda, D. (2010) Single-molecule FRET of protein-nucleic acid and protein-protein complexes: surface passivation and immobilization. *Methods,* **52**, 192–200.

[371] Swisher, J. F., Su, L. J., Brenowitz, M., Anderson, V. E., and Pyle, A. M. (2002) Productive folding to the native state by a group II intron ribozyme. *J. Mol. Biol.,* **315**, 297–310.

[372] Binder, M., Eberle, F., Seitz, S., Mucke, N., Huber, C. M., Kiani, N., Kaderali, L., Lohmann, V., Dalpke, A., and Bartenschlager, R. (2011) Molecular mechanism of signal perception and integration by the innate immune sensor retinoic acid-inducible gene-I (RIG-I). *J. Biol. Chem.,* **286**, 27278–27287.

[373] Hornung, V., Ellegast, J., Kim, S., Brzozka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K. K., Schlee, M., Endres, S., and Hartmann, G. (2006) 5'-triphosphate RNA is the ligand for RIG-I. *Science,* **314**, 994–997.

[374] Gack, M. U., Shin, Y. C., Joo, C. H., Urano, T., Liang, C., Sun, L., Takeuchi, O., Akira, S., Chen, Z., Inoue, S., and Jung, J. U. (2007) TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature,* **446**, 916–920.

[375] Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., and Fujita, T. (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.,* **5**, 730–737.

[376] Saito, T., Hirai, R., Loo, Y. M., Owen, D., Johnson, C. L., Sinha, S. C., Akira, S., Fujita, T., and Gale Jr., M. (2007) Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. *Proc. Natl. Acad. Sci. USA,* **104**, 582–587.

[377] Tanner, N. K. and Linder, P. (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol. Cell,* **8**, 251–262.

[378] Kossen, K. and Uhlenbeck, O. C. (1999) Cloning and biochemical characterization of Bacillus subtilis YxiN, a DEAD protein specifically activated by 23S rRNA: delineation of a novel sub-family of bacterial DEAD proteins. *Nucleic Acids Res.,* **27**, 3811–3820.

[379] Kossen, K., Karginov, F. V., and Uhlenbeck, O. C. (2002) The carboxy-terminal domain of the DEXD/H protein YxiN is sufficient to confer specificity for 23S rRNA. *J. Mol. Biol.,* **324**, 625–636.

[380] Ellis, R. J. and Minton, A. P. (2003) Join the crowd. *Nature,* **425**, 27–28.

[381] Sakatshii, S. K., Shubeita, G. T., and Dietler, G. (2000) On the possibility of observation of single quadrupoles by fluorescence resonance energy transfer scanning near-field optical microscopy. *Opt. Comm.,* **188**, 41–45.

[382] Selvin, P. R. (1996) Lanthanide-based resonance energy transfer. *J. Sel. Top. Quant. Electron.,* **2**, 1077–1087.

[383] Erat, M. C. and O, S. R. K. (2011) Methods to detect and characterize metal ion binding sites. *Met. Ions Life Sci.,* **9**, 37–100.

[384] Boeke, J. D. (2003) The unusual phylogenetic distribution of retrotransposons: a hypothesis. *Genome Res.,* **13**, 1975–1983.

[385] Pyle, A. M. and Marcia, M. (2012) Visualizing group II intron catalysis through the stages of splicing. *Cell,* **151**, 497–507.

[386] Brookes, P. S., Yoon, Y., Robotham, J. L., Anders, M. W., and Sheu, S. S. (2004) Calcium, ATP, and ROS: a mitochondrial love-hate triangle. *Am. J. Physiol. Cell. Physiol.,* **287**, C817–C833.

[387] Hyeon, C., Lee, J., Yoon, J., Hohng, S., and Thirumalai, D. (2012) Hidden complexity in the isomerization dynamics of Holliday junctions. *Nat. Chem.,* **4**, 907–914.

[388] Kruschel, D. and Sigel, R. K. O. (2008) Divalent metal ions promote the formation of the 5'-splice site recognition complex in a self-splicing group II intron. *J. Inorg. Chem.,* **102**, 2147–2154.

[389] Mergny, J. L. and Lacroix, L. (2003) Analysis of thermal melting curves. *Oligonucleotides,* **13**, 515–537.

[390] Kowerko, D., Hadzic, M., König, S. L. B., B"orner, R., Heidernätsch, M., and Sigel, R. K. O. Multifunctional analysis software for heterogeneous single molecule FRET data. *Manuscript in preparation,*.

[391] Chung, S. H. and Kennedy, R. A. (1991) Forward-backward non-linear filtering technique for extracting small biological signals from noise. *J. Neurosc. Methods,* **40**, 71–86.

[392] Haran, G. (2004) Noise reduction in single-molecule fluorescence trajectories of folding proteins. *Chem. Phys.,* **307**, 137–145.

[393] Sabanayagam, C. R., Eid, J. S., and Meller, A. (2005) Long time scale blinking kinetics of cyanine fluorophores conjugated to DNA and its effect on Förster resonance energy transfer. *J. Chem. Phys.,* **123**, 224708.

[394] Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez Jr., R. L., and Wiggins, C. H. (2009) Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.,* **97**, 3196–3205.

[395] Hellriegel, C., Kirstein, J., Bräuchle, C., Latour, V., Pigot, T., Olivier, R., Lacombe, S., Brown, R., Guieu, V., Payrastre, C., Izquierdo, A., and Mocho, P. (2004) Diffusion of single streptocyanine molecules in the nanoporous network of sol-gel glasses. *J. Phys. Chem. B,* **108**, 14699–14709.

[396] Flomenbom, O., Velonia, K., Loos, D., Sadahiro, M., Cotlet, M., Engelborghs, Y., Hofkens, J., Rowan, A. E., Nolte, R. J. M., van der Auweraer, M., de Schryver, F. C., and Klafter, J. (2005) Stretched exponential decay and correlations in the catalytic activity of fluctuating single lipase molecules. *Proc. Natl. Acad. Sci. USA,* **102**, 2368–2372.

[397] Greenfield, M., Pavlichin, D. S., Mabuchi, H., and Herschlag, D. (2012) Single molecule analysis research tool (SMART): an integrated approach for analyzing single molecule data. *PloS ONE,* **7**, e29703.

[398] Floyd, D. L., Harrison, S. C., and van Oijen, A. M. (2010) Analysis of kinetic intermediates in single-particle dwell time distributions. *Biophys. J.,* **99**, 360–366.

[399] Coelho, C. A. and Mexia, J. T. (2007) On the distribution of the product and ratio of independent generalized gamma-ratio random variables. *Sankhya Ser.,* **69**, 221–255.

[400] Krause, S., Kowerko, D., Börner, R., Hübner, C. G., and von Borczyskowski (2011) Spectral diffusion of single molecules in a hierarchical energy landscape. *ChemPhysChem,* **12**, 303–312.

[401] Britton, N. F. (2003) Essential mathematical biology, 1st edition, Springer, Berlin, Germany.

[402] Reed, L. J. and Berkson, J. (1929) The application of the logistic function to experimental data. *J. Phys. Chem.,* **33**, 760–779.

[403] Mitchell, P. R. (1979) Hydrophobic interactions in ternary Zinc(II) and Copper(II) complexes containing 1,10-phenanthroline or 2,2'-bipyridyl and an alkane carboxylate or sulphonate. *Dalton Trans.,* pp. 771–776.

[404] Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.,* **14**, 51–55.

[405] Rowinska-Zyrek, M., Skilandat, M., and Sigel, R. K. O. (2013) Hexaamminecobalt(III) - probing metal ion binding sites in nucleic acids by NMR. *Z. Anorg. Allg. Chemie,* **639**, 1313–1320.

[406] Reid, S. S. and Cowan, J. A. (1990) Biostructural chemistry of magnesium ion: characterization of the weak binding sites on tRNAPhe (yeast). Implications for conformational change and activity. *Biochemistry,* **29**, 6025–6032.

[407] Cowan, J. A. (1991) Coordination chemistry of $Mg^{2+}$ and 5S rRNA (*Escherichia coli*): binding parameters, ligand symmetry, and implications for activity. *J. Am. Chem. Soc.,* **113**, 675–676.

[408] Lee, J. Y., Okumus, B., Kim, D. S., and Ha, T. (2005) Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. USA,* **102**, 18938–18943.

[409] English, B. P., Min, W., van Oijen, A. M., Lee, K. T., Luo, G., Sun, H., Cherayil, B. J., Kou, S. C., and Xie, X. S. (2006) Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat. Chem. Biol.,* **2**, 87–94.

[410] Sigel, R. K. O. and Pyle, A. M. (2007) Alternative roles for metal ions in enzyme catalysis and the implications for ribozyme chemistry. *Chem. Rev.,* **107**, 97–113.

[411] Borgia, M. B., Borgia, A., Best, R. B., Steward, A., Nettels, D., Schuler, B., and Clarke, J. (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature,* **474**, 662–665.

[412] Selvin, P. R. and Ha, T. J. (2007) Single-molecule techniques, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

[413] Kowerko, D., König, S. L. B., Skilandat, M., Kruschel, D., Cardo, L., and Sigel, R. K. O. (2013) Metal ion induced kinetic heterogeneity of the intron-exon recognition in single group II introns. *Proc. Natl. Acad. Sci. USA, submitted,*.

[414] Kapanidis, A. N. and Strick, T. (2009) Biology, one molecule at a time. *Trends Biochem. Sci.,* **34**, 234–243.

[415] Walter, N. G. and Burke, J. M. (1998) The hairpin ribozyme: structure, assembly and catalysis. *Curr. Opin. Chem. Biol.,* **2**, 24–30.

[416] Yasuda, R., Masaike, T., Adachi, K., Noji, H., Itoh, H., and Kinosita Jr., K. (2003) The ATP-waiting conformation of rotating F1-ATPase revealed by single-pair fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA,* **100**, 9314–9318.

[417] Xiao, Z. and Wedd, A. G. (2010) The challenges of determining metal-protein affinities. *Nat. Prod. Rep.,* **27**, 768–789.

[418] Cui, J., Waltman, P., Le, V. H., and Lewis, E. A. (2013) The effect of molecular crowding on the stability of Human c-MYC promoter sequence i-motif at neutral pH. *Molecules,* **18**, 12751–12767.

[419] Heddi, B. and Phan, A. T. (2011) Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.,* **133**, 9824–9833.

[420] Kapinos, L. E., Song, B., and Sigel, H. (1998) Metal ion-coordinating properties of imidazole and derivatives in aqueous solution: interrelation between complex stability and ligand basicity. *Inorg. Chim. Acta,* **280**, 50–56.

[421] Kapinos, L. E., Song, B., and Sigel, H. (1999) Acid-base and metal-ion-coordinating properties of benzimidazole derivatives (= 1,3-dideazapurines) in aqueous solution: interrelation between complex stability and ligand basicity. *Chem. Eur. J.,* **5**, 1794–1802.

[422] Bowman, J. C., Lenz, T. K., Hud, N. V., and Williams, L. D. (2012) Cations in charge: magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.,* **22**, 262–272.

[423] Irving, H. and Williams, R. J. P. (1948) Order of stability of metal complexes. *Nature,* **162**, 746–747.

[424] Sigel, H. and Griesser, R. (2005) Nucleoside 5'-triphosphates: Self-association, acid-base, and metal ion-binding properties in solution. *Chem. Soc. Rev.,* **34**, 875–900.

[425] Lukavsky, P. J., Kim, I., Otto, G. A., and Puglisi, J. D. (2003) Structure of HCV IRES domain II determined by NMR. *Nat. Struct. Biol.,* **10**, 1033–1038.

[426] Walter, N. G. (2001) Structural dynamics of catalytic RNA highlighted by fluorescence resonance energy transfer. *Methods,* **25**, 19–30.

[427] Li, W., Wu, P., Ohmichi, T., and Sugimoto, N. (2002) Characterization and thermodynamic properties of quadruplex/duplex competition. *FEBS Lett.,* **526**, 77–81.

[428] Schlegel, G., Bohnenberger, J., Potapova, I., and Mews, A. (2002) Fluorescence decay time of single semiconductor nanocrystals. *Phys. Rev. Lett.,* **88**, 137401.

[429] Quinn, G. P. and Keough, M. J. Experimental design and data analysis for biologists, Cambridge University Press, Cambridge, United Kingdom, .

[430] Ennifar, E., Walter, P., and Dumas, P. (2003) A crystallographic study of the binding of 13 metal ions to two related RNA duplexes. *Nucleic Acids Res.,* **31**, 2671–2682.

[431] Sigel, H., Tribolet, R., Malini-Balakrishnan, R., and Martin, R. B. (1987) Comparison of the stabilities of monomeric metal ion complexes formed with adenosine 5'-triphosphate (ATP) and pyrimidine-nucleoside 5'-triphosphate (CTP, UTP, TTP) and evaluation of the isomeric equilibria in the complexes of ATP and CTP. *Inorg. Chem.,* **26**, 2149–2157.

[432] Uphoff, S., Gryte, K., Evans, G., and Kapanidis, A. N. (2011) Improved temporal resolution and linked hidden Markov modeling for switchable single-molecule FRET. *ChemPhysChem,* **12**, 571–579.

[433] Schuler, B. and Hoffmann, H. (2013) Single-molecule spectroscopy of protein folding dynamics - expanding scope and timescales. *Curr. Opin. Struct. Biol.,* **23**, 36–47.

[434] Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M. A., Petsko, G. A., Karplus, M., Hübner, C. G., and Kern, D. (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature,* **450**, 838–844.

[435] McKinney, S. A., Déclais, A. C., Lilley, D. M. J., and Ha, T. (2003) Structural dynamics of individual Holliday junctions. *Nat. Struct. Biol.,* **10**, 93–97.

[436] Gopich, I. V. and Szabo, A. (2007) Single-molecule FRET with diffusion and conformational dynamics. *J. Phys. Chem. B,* **111**, 12925–12932.

[437] Kelly, D., Dillingham, M. S., Hudson, A., and Wiesner, K. (2012) A new method for inferring hidden Markov models from noisy time sequences. *PloS ONE,* **7**, e29703.

[438] Qin, F. and Li, L. (2004) Model-based fitting of single-channel dwell-time distributions. *Biophys. J.,* **87**, 1657–1671.

[439] Liu, Y., Park, J., Dahmen, K. A., Chemla, Y. R., and Ha, T. (2010) A comparative study of multivariate and univariate hidden Markov modelings in time-binned single-molecule FRET data analysis. *J. Phys. Chem. B,* **114**, 5386–5403.

[440] Taylor, J. N., Makarov, D. E., and Landes, C. F. (2010) Denoising single-molecule FRET trajectories with wavelets and Bayesian inference. *Biophys. J.,* **98**, 164–173.

[441] Taylor, J. N. and Landes, C. F. (2011) Improved resolution of complex single-molecule FRET system via wavelet shrinkage. *J. Phys. Chem. B,* **115**, 1105–1114.

[442] Lee, N. K., Kapanidis, A. N., Koh, H. R., Korlann, Y., Ho, S. O., Kim, Y., Gassman, N., Kim, S. K., and Weiss, S. (2007) Three-color alternating-laser excitation of single molecules: monitoring multiple interactions and distances. *Biophys. J.,* **92**, 303–312.

[443] Haar, A. (1910) Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.,* **69**, 331–371.

[444] Little, M. A. and Jones, N. S. (2011) Generalized methods and solvers for noise removal from piecewise constant signal. *Proc. R. Soc. A,* **467**, 3088–3114.

[445] Lee, H. C., Lin, B. L., Chang, W. H., and Tu, I. P. (2012) Toward automated denoising of single molecular Förster resonance energy data. *J. Biomed. Opt.,* **17**, 3088–3114.

[446] Yang, H. and Xie, X. S. (2002) Statistical approaches for probing single-molecule dynamics photon-by-photon. *Chem. Phys.,* **284**, 423–437.

[447] Yang, H., Luo, G., Karnchanaphanurach, P., Louie, T. M., Rech, I., Cova, S., Xun, L., and Xie, X. S. (2003) Protein conformational dynamics probed by single-molecule electron transfer. *Science,* **302**, 262–266.

[448] Eddy, S. R. (2004) What is a hidden Markov model. *Nat. Biotechnol.,* **22**, 1315–1316.

[449] Jung, S. and Dickson, R. M. (2009) Hidden Markov analysis of short single molecule intensity trajectories. *J. Phys. Chem. B,* **113**, 13886–13890.

[450] Lee, T. H. (2009) Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden Markov models. *J. Phys. Chem. B,* **113**, 11535–11542.

[451] Antonik, M., Felekyan, S., Gaiduk, A., and Seidel, C. A. M. (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distribution via photon distribution analysis. *J. Phys. Chem. B,* **110**, 6970–6978.

[452] Nir, E., Michalet, X., Hamadani, K. M., Laurence, T. A., Neuhauser, D., Kovchegov, Y., and Weiss, S. (2006) Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B,* **110**, 22103–22124.

[453] Chung, H. S., Louis, J. M., and Eaton, W. A. (2010) Distinguishing between protein dynamics and dye photophysics in single-molecule FRET experiments. *Biophys. J.,* **98**, 696–706.

[454] Gopich, I. V. and Szabo, A. (2006) Theory of the statistics of kinetic transitions with application to single-molecule enzyme catalysis. *J. Chem. Phys.,* **124**, 154712.

[455] Kalinin, S., Sisamakis, E., Magennis, S. W., Felekyan, S., and Seidel, C. A. M. (2010) On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *J. Phys. Chem. B,* **114**, 6197–6206.

[456] Efron, B. (1976) Bootstrap methods: another look at the jackknife. *Ann. Stat.,* **7**, 1–26.

[457] Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., and Stamatakis, A. (2010) How many bootstrap replicates are necessary. *J. Comput. Biol.,* **17**, 337–354.

[458] Pla, L. (2004) Bootstrap confidence intervals for the Shannon biodiversity index: a simulation study. *J. Agric. Biol. Envir. S.,* **9**, 42–56.

[459] Szoszkiewicz, R., Ainavarapu, S. R. K., Wiita, A. P., Perez-Jiminez, R., Sanchez-Ruiz, J. M., and Fernandez, J. M. (2008) Dwell time analysis of a single-molecule mechanochemical reaction. *Langmuir,* **24**, 1356–1364.

[460] Hoefling, M., Lima, N., Haenni, D., Seidel, C. A. M., Schuler, B., and Grubmüller, H. (2011) Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. *PLoS ONE,* **6**, e19791.

[461] Manly, B. F. J. Randomization, bootstrap and Monte Carlo methods in biology, Chapman and Hall/CRC, Boca Raton, FL, USA, .

[462] Okamoto, K. and Sako, Y. (2012) Variational analysis of a photon-based hidden Markov model for single-molecule FRET trajectories. *Biophys. J.,* **103**, 1315–1324.

[463] MacQueen, J. T. (1967) Some observations concerning the van't Hoff equation. *J. Chem. Educ.,* **44**, 755–756.

[464] Qin, F. (2004) Restoration of single-channel currents using the segmental k-means method based on hidden Markov modeling. *Biophys. J.,* **86**, 1488–1501.

[465] Politis, D. N. (2003) The impact of bootstrap methods on time series analysis. *Stat. Sci.,* **18**, 219–230.

[466] Okamoto, K., Sannohe, Y., Mashimo, T., Sugiyama, H., and Terzima, M. (2008) G-quadruplex structures of human telomere DNA examined by single-molecule FRET and BrG substitution. *Bioorgan. Med. Chem.,* **16**, 6873–6879.

[467] Okamoto, K. and Terzima, M. (2008) Distribution analysis for single-molecule FRET measurement. *J. Phys. Chem. B,* **112**, 7308–7314.

[468] Su, L. J., Qin, P. Z., Michels, W. J., and Pyle, A. M. (2001) Guiding ribozyme cleavage through motif recognition: the mechanism of cleavage site selection by a group II intron ribozyme. *J. Mol. Biol.,* **306**, 655–668.

[469] Kim, H. K., Rasnik, I., Juewen, L., Ha, T., and Lu, Y. (2007) Dissecting metal ion-dependent folding and catalysis of a single DNAzyme. *Nat. Chem. Biol.,* **3**, 763–768.

[470] König, S. L. B., Kowerko, D., Khier, M., and Sigel, R. K. O. Cation-dependent formation of RNA structure dissected by single-molecule fluorescence. *Nat. Chem., submitted,*.

[471] Hess, S. T., Girirajan, T. P. K., and Mason, M. D. (2006) Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.,* **91**, 4258–4272.

[472] Giudici, P., Rydén, T., and Vandekerkhove, P. (2000) Likelihood-ratio tests for hidden Markov models. *Biometrics,* **56**, 742–747.

[473] Börner, R., Kowerko, D., Krause, S., von Borczyskowsi, C., and Hübner, C. (2012) Efficient simultaneous fluorescence orientation, spectrum, and lifetime detection for single molecule dynamics. *J. Chem. Phys.,* **137**, 164202.

[474] Barlett, J. S., Voss, K. J., Sathyendranath, S., and Vodacek, A. (1998) Raman scattering by pure water and seawater. *Applied Optics,* **37**, 3324–3332.

[475] Mishra, A., Behera, R. K., Behera, P. K., Mishra, B. K., and Behera, G. B. (2000) Cyanines during the 1990s: a review. *Chem. Rev.,* **100**, 1973–2011.

[476] Taraska, J. W., Puljung, M. C., Olivier, N. B., Flynn, G. E., and Zagotta, W. N. (2009) Mapping the structure and conformational movements of proteins with transition metal ion FRET. *Nat. Methods,* **6**, 532–537.

[477] Taraska, J. W., Puljung, M. C., and Zagotta, W. N. (2009) Short-distance probes for protein backbone structure based on energy transfer between bimane and transition metal ions. *Proc. Natl. Acad. Sci. USA,* **106**, 16227–16232.

[478] Hermanson, G. T. Pegylation and synthetic polymer modification. *In*: Bioconjugate techniques pp. 787–938 Elsevier Inc, Waltham, MA, USA (2013).

[479] Morris, S. J., Bradley, D., and Blumenthal, R. (1985) The use of cobalt ions as a collisional quencher to probe surface charge and stability of fluorescently labeled bilayer vesicles. *Biochimica et Biophysica Acta,* **818**, 365–372.

[480] Homan, R. and Eisenberg, M. (1985) A fluorescence quenching technique for the measurement of paramagnetic ion concentrations at the membrane/water interface. Intrinsic and X537-mediated cobalt fluxes across lipid bilayer membranes. *Biochimica et Biophysica Acta,* **812**, 485–492.

[481] Holmes, A. S., Birch, D. J. S., Suhling, K., and Imhof, R. E. (1991) Evidence for donor-donor energy transfer in lipid bilayers: perylene fluorescence quenching by $Co^{2+}$. *Chem. Phys. Lett.,* **186**, 189–194.

[482] Salthammer, T. and Dreeskamp, H. (1990) Fluorescence quenching of perylene by $Co^{2+}$ ions via energy transfer in viscous and non-viscous media. *J. Photochem. Photobiol. A: Chem.,* **55**, 53–62.

[483] Hariharan, C., Vijaysree, V., and Mishra, A. K. (1997) Quenching of 2,5-diphenyloxazole (PPO) fluorescence by metal ions. *J. Luminesc.,* **75**, 205–211.

[484] Demchenko, A. P. (2008) Site-selective red-edge effect. *Method. Enzymol.,* **450**, 49–78.

[485] Thompson, R. B., Ge, Z., Patchan, M. W., and Fierke, C. A. (1996) Performance enhancement of fluorescence energy transfer-based biosensors by site-directed mutagenesis of the transducer. *J. Biomed. Opt.,* **1**, 131–137.

[486] Bagshaw, C. R. and Cherny, D. (2006) Blinking fluorophores: what do they tell us about protein dynamics. *Biochem. Soc. Trans.,* **34**, 515–527.

# Chapter 10

# Summary

It is increasingly clear that RNAs are involved in numerous aspects of cellular metabolism. However, an RNA molecule achieves functionality solely upon adopting a precise three-dimensional fold, which consists of an arrangement of secondary and tertiary structure elements. To form such higher-order structures, the electrostatic repulsion created by the negative charges along the phosphate-sugar backbone has to be overcome by metal ions. An estimated 10-20 % of these metal ions interact with RNA in a highly specific manner, both with regard to the identity of the cation and its binding site. Due to its relatively high abundance inside the cell and its propensity of coordinate to oxygen, $Mg^{2+}$ is generally believed to be the most important metal ion cofactor for RNA folding and catalysis. However, a considerable body of literature has shown that numerous RNAs respond to the presence of different cations by altering their fold, their catalytic activity or even their catalytic repertoire. In these cases, one may speculate that $Mg^{2+}$ is not the natural cofactor and/or the RNA may be controlled via a metal ion switch. Albeit, the lack of systematic studies addressing cation-mediated RNA folding and catalysis precludes solid interpretation.

One large RNA enzyme (ribozyme) is the yeast group IIB intron *Sc*.ai5$\gamma$, which catalyses its own excision from precursor messenger RNA, followed by joining the adjacent coding regions (exons) together. *Sc*.ai5$\gamma$ consists of six functionally distinct domains that interact via long-range tertiary contacts (Figure 10.1A). Site specificity of the cleavage reaction is ensured by proper base pairing between distal exon-binding sites (EBS) and their cognate intron binding sites (IBS). Intron folding and catalysis is dependent on $Mg^{2+}$ and is perturbed by small amounts of $Ca^{2+}$. Even though group II introns have not (yet) been identified in the nuclear genomes of eukaryotes, there is accumulating evidence that they are ancestors of the spliceosome and retrotransposons. As spliceosomal introns and retrotransponsons comprise more than half of the human genome, group II introns have received considerable attention in recent years.

Single-molecule Förster resonance energy transfer (smFRET) reports on the extent of non-radiative energy transfer between a single pair of fluorophores, which allows to estimate their relative distances. When fluorophores are conjugated to a biomolecule, smFRET can thus be applied as a “molecular ruler” reporting on macromolecular structure and dynamics. This permits to study the full folding pathway of the biomolecule, unveiling conformational intermediates and intermolecular heterogeneity. This is an important feature of single-molecule spectroscopy, because the vast conformational space sampled by biopolymers often leads to the occurrence of kinetic traps and/or multiple native states that remain hidden when conventional ensemble methods are used. The enormous potential of single-molecule approaches in biological research is further illustrated in the review presented in Chapter 4.

This work sets out to characterise metal ion dependent formation of RNA tertiary structure using single-molecule FRET. For this purpose, a previously validated pair of functional oligonucleotides derived from the 5’ exon-intron recognition site in the Sc.ai5$\gamma$ ribozyme (d3’EBS1* and IBS1*) was labelled with the FRET pair Cy3 and Cy5 (Figure 10.1B). d3’EBS1* was further conjugated to a biotin moiety, which allows for immobilisation on a surface and imaging of single fluorophore emission over several minutes using a total internal reflection fluorescence microscope. Thus, cation-dependent docking/undocking dynamics of single d3’EBS1* and IBS1* molecules could be monitored in real time.

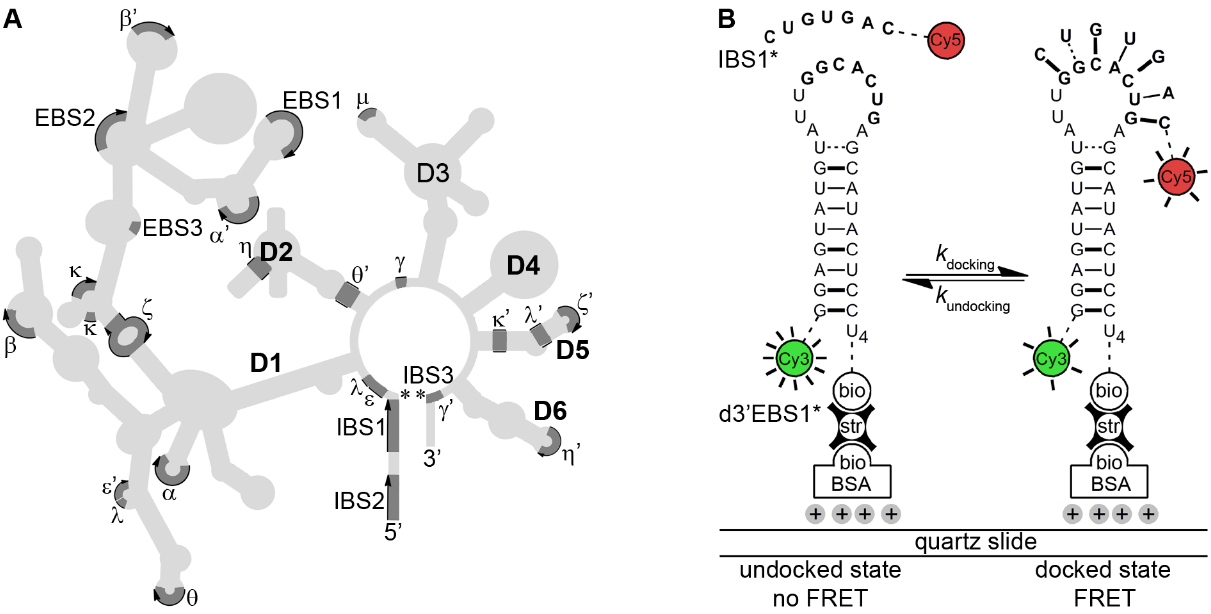


Figure 10.1: Secondary structure of the yeast group IIB intron ai5$\gamma$ and experimental design. (A) Six domains (D1, ..., D6) are arranged around a central wheel and achieve the active fold upon formation of numerous long-range tertiary interactions (Greek letter code). Exon- and intron-binding sites (EBS, IBS) are required for splice site (*) recognition. (B) Experimental design to study 5’ splice site formation by single-molecule FRET. The d3’EBS1* hairpin is Cy3-labelled and tethered to the surface of a quartz slide via a biotin-streptavidin linkage. Docking of IBS1* is accompanied by the appearance of Cy5 fluorescence and a concomitant decrease of Cy3 emission due to FRET.

**Metal ion induced kinetic heterogeneity of the intron-exon recognition in single group II introns**

This Chapter addresses the influence of the physiologically important cations $Mg^{2+}$ and $Ca^{2+}$ on d3'EBS1*/IBS1* interaction using an unprecedented combination of single-molecule FRET and NMR. Single-molecule *FRET* time traces reveal two well separated *FRET* states centered around 0 (d3'EBS1*) and ∼0.7 (d3'EBS1*/IBS1*, Figure 10.2A).1D *FRET* histograms created from several hundred single-molecule time traces clearly demonstrate that the presence of 100 mM $K^+$ is not sufficient for stable interstrand association (Figure 10.2B). In turn, addition of physiological amounts of $Mg^{2+}$ or $Ca^{2+}$ leads to a pronounced shift of the thermodynamic equilibrium towards the docked state, a finding that was confirmed by ensemble spectroscopy (Figure 10.2C).

Visual inspection of time traces reveals pronounced differences between single molecules that were imaged side-by-side under identical conditions, both with regard to interconversion rates and relative occupation of the two FRET states (Figure 10.2A). Kinetic analysis of time traces recorded in the absence of $M^{2+}$ demonstrates that IBS1* docking can be described with a single rate constant, while undocking is a (slightly) heterogeneous process that requires three rate constants to be satisfactorily described. Addition of divalent metal ions leads to the appearance of heterogeneity in docking and increases the heterogeneity associated with the undocking process. Using Monte-Carlo simulations, it is shown that finite observation time can only account for a small fraction of this heterogeneity, followed by providing a mathematical framework to quantify the effect in $K_A$ distributions that are broadened due to pronounced intermolecular heterogeneity. The occurrence of kinetic heterogeneity suggests a rugged free energy landscape with more than one thermodynamically stable d3'EBS1*/IBS1* conformation. $M^{2+}$ appears to further stabilise (local) energy minima, leading to substructures with similar global folds, but differences on the microscopic level.

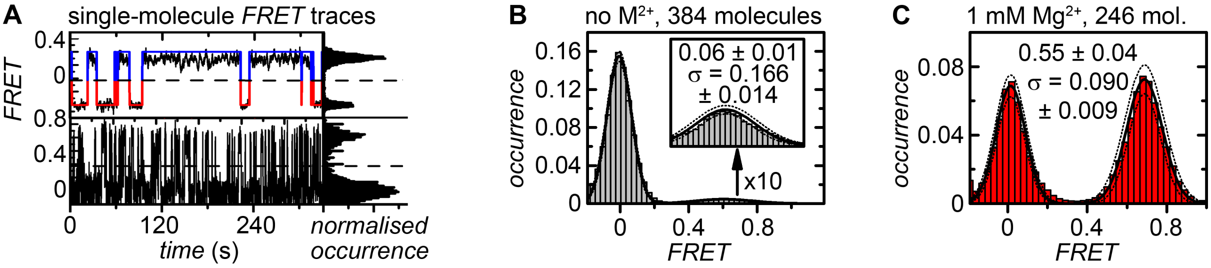


Figure 10.2: Studying cation-dependent interaction of d3'EBS1* and IBS1* by single-molecule fluorescence. (A) Representative *FRET* time traces reveal fluctuations between the docked and the undocked conformation (high and zero *FRET*). The red/blue line corresponds to the discretised profile, the dashed lines represent an arbitrary threshold at *FRET* = 0.375 to distinguish the two states. Both traces were recorded at 100 mM $K^+$ and 8 mM $Ca^{2+}$. (B, C) Normalised cumulated *FRET* histograms built from single time traces recorded at 100 mM $K^+$ or 100 mM $K^+$ and 1 mM $Mg^{2+}$, respectively. Solid lines correspond to Gaussian fits, dashed lines visualise the cross-sample variability as quantified by bootstrapping ($2\sigma_{\text{bootstrap}}$). The inset shows a 10-fold magnification of the high *FRET* distribution.

In order to rationalise these heterogeneous kinetics, we turned our attention to the solution structures of d3'EBS1* and d3'EBS1*/IBS1*, in particular the specific $Mg^{2+}$ binding sites (Figure 10.3A). NMR titration studies were used to pinpoint the nature of $Mg^{2+}$ binding, suggesting that the two metal ions mediating EBS1*/IBS1* interaction are partly dehydrated. Such inner-sphere coordination is expected to strongly decrease the cation exchange rate. Estimation of the affinity of the metal ions towards their cognate binding sites further suggests that the binding sites are likely to be only partly occupied at physiological cation concentrations (Figure 10.3A).

Taken together, the combined single-molecule FRET and NMR data demonstrate that intermolecular heterogeneity coincides with incomplete occupation of metal ion binding sites. Conformational changes in RNA have often been observed to display multiexponential kinetics, but discord persists with regard to the origin of the phenomenon (see also: Chapter 6). Here, we show that microscopically different substructures are likely to the be caused by incomplete/heterogeneous occupation of metal ion binding sites, providing an explanation for a long-standing question in the field of nucleic acids biophysics (Figure 10.3B).

### Cation-promoted RNA tertiary structure formation dissected by single-molecule fluorescence

Metal ions interacting with RNA typically bind to negatively charged phosphate moieties along the phosphate-sugar backbone and to exo- and endocyclic heteroatoms within the purine/pyrimidine nucleobases (Figure 10.4A). The affinity of divalent metal ions towards these binding

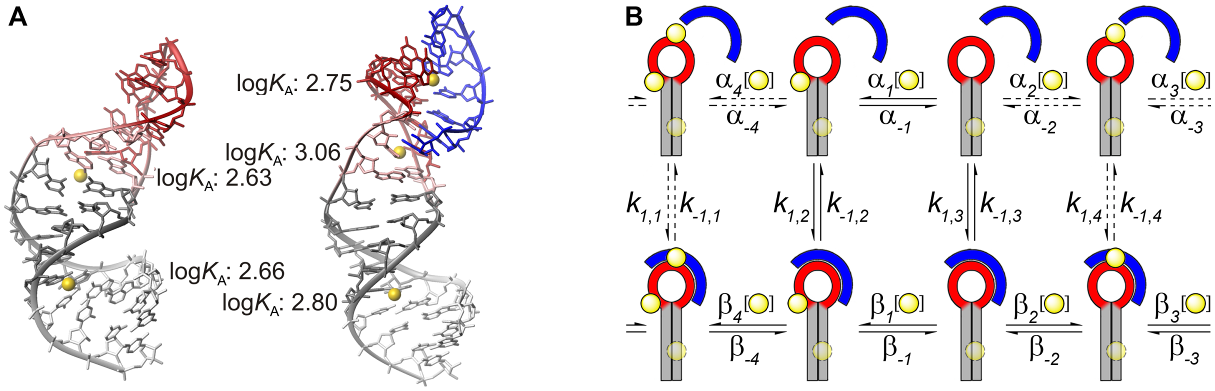


Figure 10.3: (A) Solution structure and proposed metal ion binding sites of d3'EBS1* and d3'EBS1*/IBS1*. The lowest energy structures from an ensemble of 20 are shown. Yellow spheres mark possible binding positions of $Mg^{2+}$ for which an estimate of the affinity constant is given. (B) Working model for cation-dependent heterogeneity in d3'EBS1*/IBS1* interaction. Several d3'EBS1* conformers differing in metal ion binding site occupation co-exist at physiological $Mg^{2+}$ or $Ca^{2+}$ concentrations and they interconvert at the cation exchange rates ($\alpha$, $\beta$). Conformers further differ with regard to IBS1* docking and undocking kinetics ($k_1$, $k_{-1}$). The distant cation binding site located near the G4-C26 base pair is assumed to have little influence on docking/undocking rates. Dashed lines indicate conversions that were not detected experimentally.

sites has been studied in isolation using small molecule mimics (Figure 10.4B). Here, the stability of acetate, imidazole, and benzimidazole complexes follows the order $Ba^{2+} < Sr^{2+} < Ca^{2+} < Mg^{2+} < Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} << Cu^{2+} > Zn^{2+} < Cd^{2+}$, a trend that is commonly known as Irving-Williams series. In turn, the Irving-Williams series is broken when the metal ion coordinates to phosphate ligands, as these complexes display a local maximum in stability in the presence of $Mn^{2+}$ (Figure 10.4B). This Chapter describes the first application of single-molecule FRET to disentangle the influence of divalent metal ions along the extended Irving-Williams series on the formation of RNA tertiary structure.

Single-molecule FRET experiments were performed in the presence of 100 mM $K^+$ and 1 mM $M^{2+}$ (M = Ba, Sr, Ca, Mg, Mn, Co, Ni or Cd). Cumulated *FRET* histograms reveal that divalent metal ions strongly differ with regard to their ability to mediate stable interstrand association, a trend that was confirmed in UV thermal melting experiments. When the affinity between d3'EBS1* and IBS1* is plotted along the extended Irving-Williams series (Figure 10.4A), it becomes evident that stability reaches a local maximum in the presence of $Mn^{2+}$. Thermodynamic stability decreases in the presence of $Co^{2+}$ and $Ni^{2+}$, followed by reaching very high values in the presence of $Zn^{2+}$ and $Cd^{2+}$ as quantified by UV thermal melting experiments. smFRET data recorded in the presence of $Zn^{2+}$ and $Cd^{2+}$ are erroneous because of solubility and surface passivation issues, as well as interaction with the surface passivation agent BSA.

Analysis of kinetics was performed to further assess the influence of the metal ion cofactor on docking and undocking. Here, dwell time histograms were fitted to stretched exponential decay models accounting for the cation-induced intermolecular heterogeneity described in the

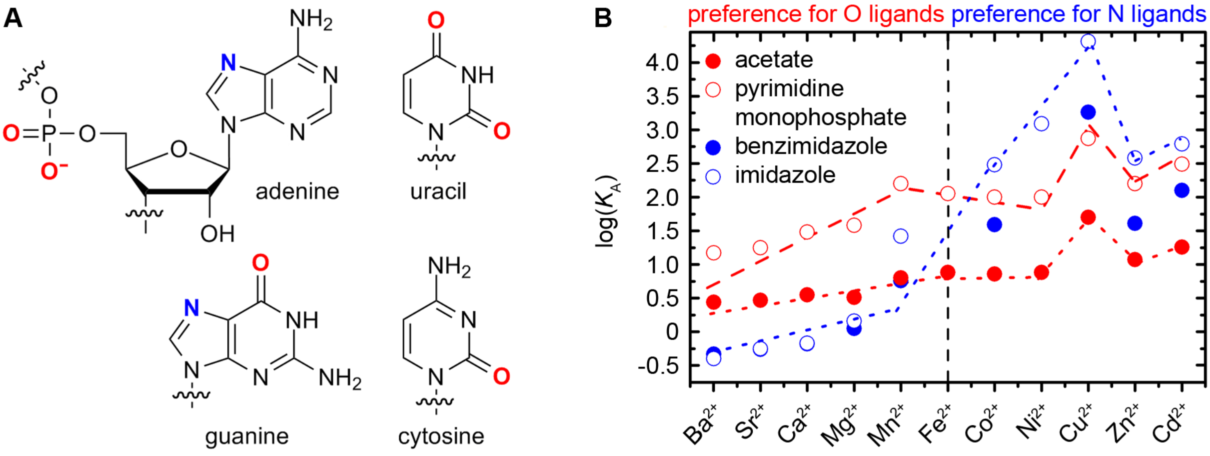


Figure 10.4: Metal ion coordination to the RNA backbone and ribonucleobases. (A) Schematic overview of the major cation binding sites located within the building blocks of RNA (coloured atoms). (B) Stability of metal ion complexes using ligands mimicking metal ion binding sites within RNA. While the stability of acetate, benzimidazole, and imidazole complexes follow the Irving-Williams series, a local maximum is observed for manganese phosphate monoester complexes. Lines represent a visual guide to the data.

previous section. The rate of docking ($k_{\text{docking}}$) is found to steadily increase along the extended Irving-Williams series and reaches a maximum in the presence of $Ni^{2+}$ (Figure 10.4B). In turn, $k_{\text{undocking}}$ reaches a local minimum in the presence of $Mn^{2+}$, followed by an increase in the presence of $Co^{2+}$ and $Ni^{2+}$ (Figure 10.4B). These data clearly demonstrate that the metal ion cofactor impacts both the docking and the undocking reaction. It should be emphasised, however, that the relative differences in $k_{\text{docking}}$ do not exceed a factor of 2, whereas cation-dependent undocking rates may differ by one order of magnitude. As a consequence, differences in thermodynamic stability of the docked d3'EBS1*/IBS1* complex can for the most part be attributed to cation-specific stabilisation of the docked state and not an acceleration of the docking process.

Comparison of the combined experimental data (Figure 10.3) with the coordination properties of the metal ions assessed herein (Figure 10.4) reveals that the thermodynamic stability of d3'EBS1*/IBS1* is governed by the affinity of the cofactor towards phosphate ligands. This provides strong evidence that the divalent metal ions mediate stable RNA-RNA interaction by specific coordination to phosphate moeities. In turn, the rate of docking follows the classical Irving-Williams series, which implies that the propensity of the cation to coordinate to phosphate ligands does not significantly influence the rate of docking. Rather, docking is likely to rely on non-specific charge screening along the RNA. In conclusion, this study not only validates smFRET as a tool to systematically study cation-mediated RNA structure formation, it also demonstrates that RNA-RNA duplex formation can be modulated according to the intrinsic coordination chemistry of the metal ion cofactor.

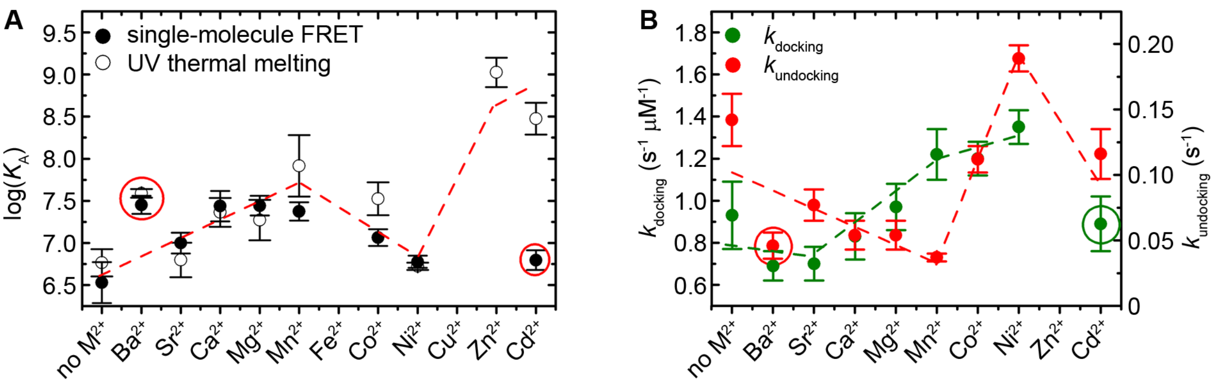
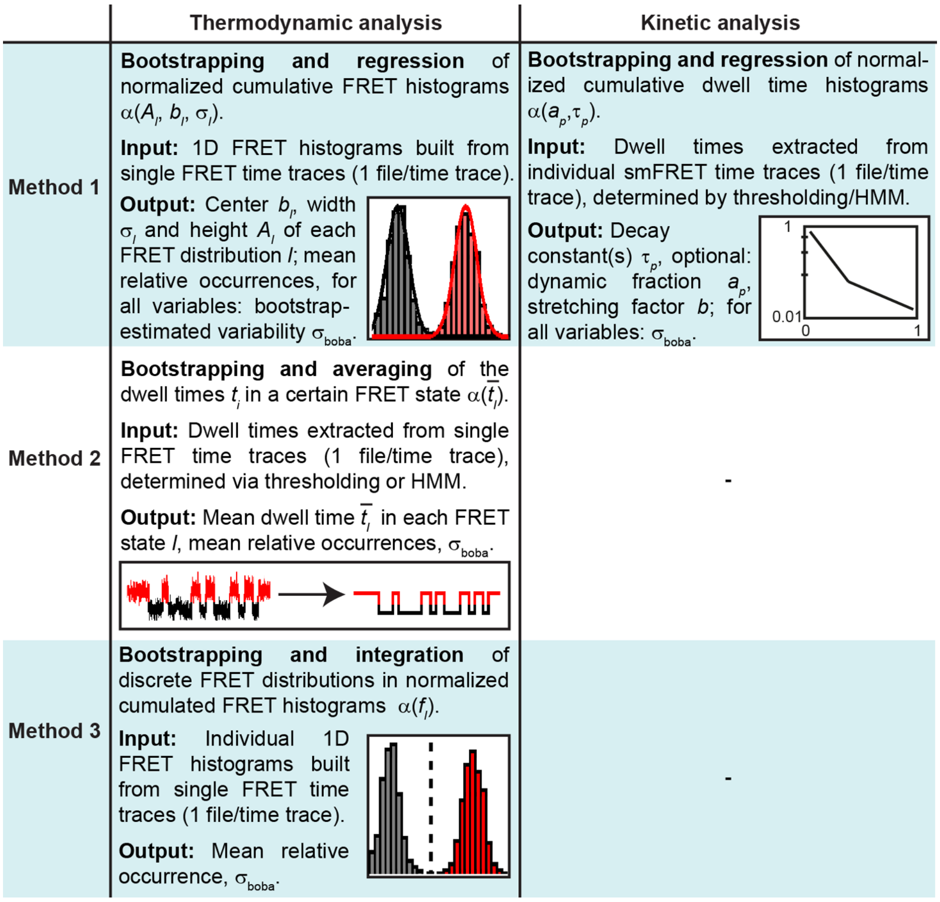


Figure 10.5: Cation-dependent thermodynamics and kinetics of d3'EBS1*/IBS1* interaction. (A) Stability constants of the docked d3'EBS1*/IBS1* complex in the presence of 100 mM $K^+$ and 1 mM $M^{2+}$ as determined by single-molecule FRET and UV thermal melting experiments. (B) Kinetics of IBS1* docking/undocking in the presence of different divalent cations. Kinetic rate constants were determined by fitting dwell time histograms to stretched exponential models. The variability associated with smFRET experiments was estimated by bootstrapping ($3\sigma_{\text{bootstrap}}$), while the error bars associated with UV melting data were determined from biological replicates ($1\sigma$). Lines represent a visual guide to the data.

### Bootstrap-based analysis of single-molecule FRET data

As smFRET permits to follow folding and catalysis of single molecules in real-time, it has become an increasingly popular method in recent years. However, individual molecules usually display pronounced differences in the relative population of states and the corresponding conversion rates, an observation that is due to limited observation time and/or unconventional kinetics (see also: Chapter 5 and 6). Quantification of such cross-sample variability is necessary to perform statistical testing in order to assess whether changes observed in response to an experimental parameter are significant. However, both estimation of cross-sample variability and statistical hypothesis testing have been disregarded to date, precluding robust biological interpretation. This Chapter provides the mathematical framework to estimate cross-sample variability in time-binned smFRET experiments by bootstrapping, followed by statistical hypothesis testing.

Bootstrapping is a resampling method that uses a sample distribution of the size $n$ to compute the distribution of the whole population, *i.e.* it provides an estimation of the mean and the



| | Thermodynamic analysis | Kinetic analysis |
|---|---|---|
| **Method 1** | **Bootstrapping and regression** of normalized cumulative FRET histograms $\alpha(A_l, b_l, \sigma_l)$.<br>**Input:** 1D FRET histograms built from single FRET time traces (1 file/time trace).<br>**Output:** Center $b_l$, width $\sigma_l$ and height $A_l$ of each FRET distribution $l$; mean relative occurrences, for all variables: bootstrap-estimated variability $\sigma_{boba}$. | **Bootstrapping and regression** of normalized cumulative dwell time histograms $\alpha(a_p,\tau_p)$.<br>**Input:** Dwell times extracted from individual smFRET time traces (1 file/time trace), determined by thresholding/HMM.<br>**Output:** Decay constant(s) $\tau_p$, optional: dynamic fraction $a_p$, stretching factor $b$; for all variables: $\sigma_{boba}$. |
| **Method 2** | **Bootstrapping and averaging** of the dwell times $t_l$ in a certain FRET state $\alpha(\bar{t}_l)$.<br>**Input:** Dwell times extracted from single FRET time traces (1 file/time trace), determined via thresholding or HMM.<br>**Output:** Mean dwell time $\bar{t}_l$ in each FRET state $l$, mean relative occurrences, $\sigma_{boba}$. | - |
| **Method 3** | **Bootstrapping and integration** of discrete FRET distributions in normalized cumulated FRET histograms $\alpha(f_l)$.<br>**Input:** Individual 1D FRET histograms built from single FRET time traces (1 file/time trace).<br>**Output:** Mean relative occurrence, $\sigma_{boba}$. | - |

Figure 10.6: Summary of the different functionalities implemented in BOBA FRET. The respective input and output variables are indicated as well. Please refer to Chapter 8 for a detailled mathematical description.

variance. It is demonstrated that the bootstrapping scheme is fully compatible with commonly used approaches analyse time-binned smFRET data (Figure 10.6). Subsequently, simulated time traces are used to assess the robustness of bootstrapping performed in conjunction with six different methods commonly employed to extract thermodynamic and kinetic parameters. The results of this computational study demonstrate that bootstrap-estimated means and cross-sample variabilities faithfully cover theoretical values, unless the approch chosen to analyse *FRET* histograms or extract dwell times is inappropriate and leads to biased input values. Finally, a set of experimental data is analysed by bootstrapping and it is shown that divalent metal metal ions have a significant effect on both thermodynamic and kinetic aspects of d3'EBS1*/IBS1* interaction.

Taken together, this study validates the bootstrap scheme for the quantification cross-sample variability in smFRET studies. Additionally, a freely available and easily implementable algorithm with graphical user interface for bootstrap-based analysis of smFRET data is provided. As estimation of cross-sample variability provides a handle on robust biological interpretation of smFRET data, bootstrapping is anticipated to be widely adapted in the field of single-molecule FRET.

# Chapter 11

# Zusammenfassung

Im Laufe der letzten Jahre hat sich die Bedeutung vom Ribonukleinsäuren (RNS) für den Stoffwechsel der Zelle immer deutlicher abgezeichnet. Die Funktionalität eines RNS-Moleküls setzt eine entsprechende dreidimensionale Struktur voraus, die aus einer genau definierten Anordnung von Sekundär- und Tertiärstrukturelementen besteht. Aufgrund der negativen Ladungen entlang des Rückrades der RNS hängt die Stabilität solcher Strukturen direkt von der Wechselwirkung mit positiv geladener Metallionen ab. Es wird geschätzt, dass 10-20 % dieser Metallionen hochspezifisch mit der RNS interagieren, d.h. das RNS-Molekül besitzt Bindungstaschen, in denen nur bestimmte Metallionen stabil gebunden werden können. Aufgrund relativ hoher intrazellulärer Konzentration und der ausgeprägten Tendenz, an Sauerstoff-Liganden zu binden, wird im Allgemeinen angenommen, dass $Mg^{2+}$ in biologischen Systemen den wichtigsten Metallionen-Cofaktor darstellt. An dieser Stelle muss darauf hingewiesen werden, dass zahlreiche Studien anhand diverser RNS-Systeme gezeigt haben, dass sich durch Austausch von $Mg^{2+}$ durch andere Kationen die Faltung, die katalytische Aktivität oder sogar das katalytische Repertoire eines RNS-Moleküls ändern kann. In diesem Fall stellt $Mg^{2+}$ möglicherweise nicht den natürlichen Cofaktor dar und/oder die Zelle reguliert die Funktion der RNS mittels Metallionen. Solche Annahmen sind jedoch aufgrund der Seltenheit systematischer Untersuchungen des Einflusses verschiedener Kationen auf die Faltung und die katalytische Aktivität von RNS spekulativer Natur.

RNS-Moleküle, die als biologische Katalysatoren fungieren, werden als Ribozyme bezeichnet. Das in den Mitochondrien von *S. cerevisiae* vorkommende Gruppe-II-Intron *Sc*.ai5$\gamma$ ist ein solches Ribozym. Es schneidet sich aus der prä-mRNS heraus und katalysiert anschließend die Ligation der benachbarten Exons. *Sc*.ai5$\gamma$ besteht aus sechs Faltdomänen, denen in der Katalyse jeweils unterschiedliche Funktionen zukommen (Abbildung 11.1A). Die einzelen Faltdomänen interagieren hierbei durch distale Tertiärkontakte. So gewährleistet beispielweise die Wechselwirkung von Exon- und Intron-Bindungsstellen (EBS, IBS) die korrekte Erkennung der Spleißstellen. Zudem muss noch darauf hingewiesen werden, dass die katalytische Aktivität von Gruppe-II-Introns stark $Mg^{2+}$-abhängig ist und schon durch kleine Mengen $Ca^{2+}$ inhibiert wird. Obgleich Gruppe-II-Introns (noch) nicht in den nuklearen Genomen höherer Eukaryoten gefunden wurden, deutet Vieles darauf hin, dass sowohl das Spliceosom als auch Retrotransposons von ihnen abstammen. Angesichts der Tatsache, dass durch das Spliceosom entfernte, nicht-codierende Sequenzen und Retrotransposons mehr als die Hälfte des menschlichen Genoms ausmachen, hat das Interesse an Gruppe-II-Introns in den vergangenen Jahren stetig zugenommen.

Förster-Resonanzenergietransfer (FRET) ist ein physikalischer Prozess, bei dem die Energie eines elektronisch angeregten Farbstoffs (Donor) auf einen zweiten Farbstoff (Akzeptor) übertragen wird. Da die Effizienz dieses Prozesses vom Abstand der Fluorophore abhängt, kann ein solches FRET-Paar als "molekularer Distanzmesser" eingesetzt werden. Das Messen von FRET zwischen einzelnen an (Bio-)Moleküle konjugierten Farbstoffpaaren wird als Einzelmolekül-FRET (*engl.* "smFRET") bezeichnet und dient der Aufklärung biomolekularer Faltungswege. Die Stärke der Einzelmolekülspektroskopie beruht auf der Tatsache, dass nicht über das Molekülensemble gemittelt wird und somit sowohl Faltungsintermediate als auch intermolekulare Unterschiede detektiert werden. Aufgrund der nahezu unbegrenzten Faltungsmöglichkeiten eines Biomoleküls sind Fehlfaltungen und alternative aktive Strukturen keine Seltenheit. Für eine Ausführung des Potentials von Einzelmokülmethoden sei auf Kapitel 4 verwiesen.

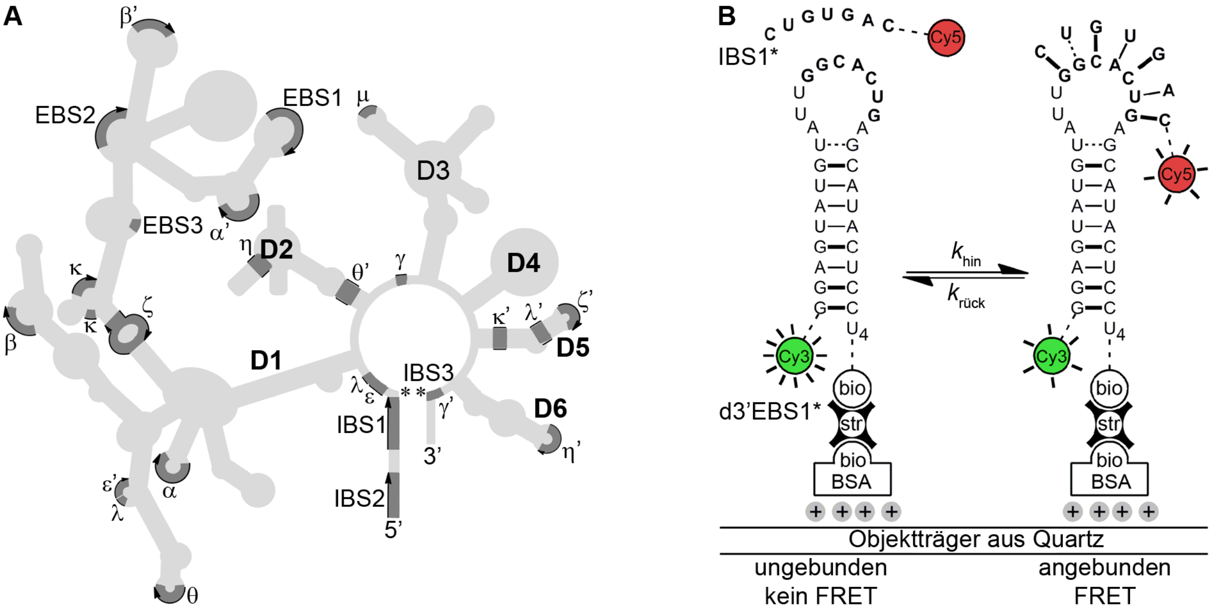


Figure 11.1: Sekundärstruktur des Gruppe-IIB-Introns ai5$\gamma$ und Planung der Einzelmolekülexperimente. (A) ai5$\gamma$ besteht aus sechs kreisförmig angeordneten Faltdomänen, die nach der Ausbildung zahlreicher Tertiärkontakte (griechischer Buchstabencode) die native Konformation annehmen. Exon- und Intron-Bindungsstellen (EBS, IBS) gewährleisten dabei die präzise Erkennung der Spleißstelle (*). (B) Untersuchung der Erkennung des 5'-Spleißstelle mittels smFRET. Die d3'EBS1* Stammschleifenstruktur ist mit einem Cy3-Farbstoff versehen und mittels einer Biotin-Streptavidin-Bindung auf der Oberfläche eines Objektträgers immobilisiert. Die Anbindung eines IBS1*-Moleküls ist mit dem Auftreten von Cy5-Fluoreszenz und einem Abfall von Cy3-Fluoreszenz aufgrund von FRET verbunden.

Das wissenschaftliche Ziel dieser Arbeit ist die Untersuchung des Einflusses von Metallionen auf die Ausbildung von RNS-Tertiärkontakten mittels smFRET. Als Modellsystem wurden hierfür zwei funktionelle RNS-Sequenzen aus dem Gruppe-II-Intron Sc.ai5$\gamma$ herangezogen (d3'EBS1* und IBS1*), die in der Zelle die Erkennung der 5'-Spleißstelle gewährleisten. d3'EBS1* wurde am 5'-Ende mit dem Farbstoff Cy3 und am 3'-Ende mit einer Biotin-Gruppe versehen. Die so markierten Oligonukleotide wurden auf der Oberfläche eines Quartz-Objektträgers immobilisert und die Fluoreszenz einzelner Farbstoffe wurde mit Hilfe eines Totalreflexionsfluoreszenzmikroskops über eine Dauer von mehreren Minuten gemessen. Anbindung und Dissoziation in Lösung befindlicher, mit Cy5-markierter IBS1*-Moleküle konnte dadurch in Echtzeit und mit Einzelmolekülauflösung detektiert werden (Abbildung 11.1B).

**Kinetische Heterogeneität in der Erkennung der 5'-Spleißstelle in Gruppe-II-Intronsn**

In diesem Kapitel wird mittels einer neuartigen Kombination von smFRET und NMR der Einfluss der für die Zelle essentiellen Kationen $Mg^{2+}$ und $Ca^{2+}$ auf die Wechselwirkung von d3'EBS1* und IBS1* beschrieben. Repräsentative Einzelmolekülzeitreihen fluktuieren zwischen einer *FRET*-Effizienz von 0 (d3'EBS1*) und ~0.7 (d3'EBS1*/IBS1*, Abbildung 11.2A). Aus über 200 Zeitreihen erstellte *FRET*-Histogramme zeigen zudem, dass eine $K^+$-Konzentration von 100 mM für eine stabile Anbindung von IBS1* nicht ausreichend ist (Abbildung 11.2B). Zugabe physiologischer Mengen an $Mg^{2+}$ oder $Ca^{2+}$ verschieben das Gleichgewicht stark in Richtung des d3'EBS1*/IBS1*-Komplexes (Figure 10.2C). Der ausgeprägte Einfluss divalenter Metallionen auf das Wechselwirkungsverhalten der beiden Oligonukleotide wurde mit Hilfe von Ensemblespektroskopie bestätigt.

Der Vergleich unter identischen Bedingungen aufgenommener Zeitreihen macht deutlich,

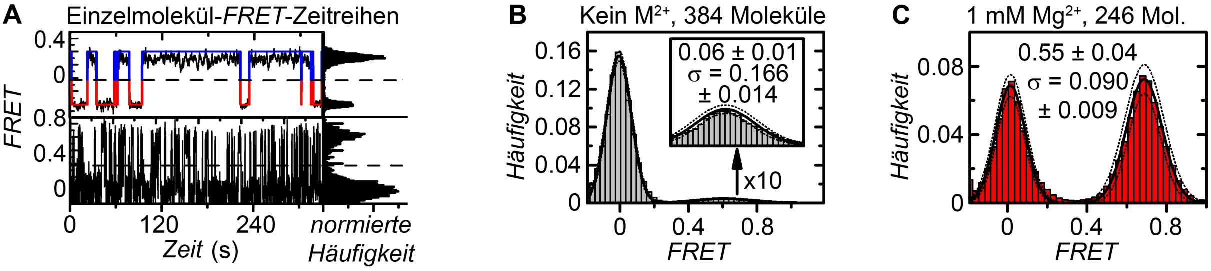


Figure 11.2: Untersuchung der kationenabhängingen Wechselwirkung von d3'EBS1* und IBS1* mittels Einzelmolekülspektroskopie. (A) Repräsentative *FRET*-Zeitreihen. Der *FRET*-Wert fluktuiert zeitabhängig zwischen 0 (gebundener Zustand) und ~0.7 (gebundener Zustand). Die blau-rote Linie stellt eine mittels Schwellwertanalyse (Schwellwert = 0.375) durchgeführte Idealisierung der experimentellen Daten dar. Beide Zeitreihen wurden in 100 mM $K^+$ und 8 mM $Ca^{2+}$ aufgenommen. (B, C) Normierte kumulierte *FRET*-Histogramme. Zeitspuren wurden in 100 mM $K^+$ bzw. 100 mM $K^+$ und 1 mM $Mg^{2+}$ aufgenommen, histogrammiert, kumuliert und anschließend normiert. Die durchgezogenen Linien entsprechen an die experimentellen Daten angepassten Normalverteilungen. Gestrichelte Linien zeigen die durch Randomisierung abgeschätzte Varianz an ($2\sigma_{\text{bootstrap}}$).

dass sich einzelne Zeitspuren hinsichtlich der im Mittel im gebundenen Zustand verbrachten Zeit und hinsichtlich der Geschwindigkeitskonstanten $k_{\text{hin}}$ und $k_{\text{rück}}$ unterscheiden müssen (Abbildung 11.2A). Kinetische Untersuchungen in Abwesenheit von $M^{2+}$ aufgenommener Zeitreihen ergibt, dass die Anbindung von IBS1* mit einer Geschwindigkeitskonstante beschrieben werden kann. Im Gegensatz dazu handelt es sich bei der Dissoziation um einen heterogenen Prozess, für dessen mathematische Beschreibung drei (oder mehr) Geschwindigkeitskonstanten benötigt werden. Zugabe divalenter Metallionen führt zum Auftreten kinetischer Heterogenität in der Anbindungsreaktion und verstärkt gleichzeitig die mit der Dissoziation verbundene Heterogeneität. Mittels Monte-Carlo-Simulationen wird gezeigt, dass die begrenzte Messdauer in Einzelmolekükexperimenten diese Beobachtung nicht ausreichend erklären kann. Anschließend wird die durch kinetische Heterogenität hervorgerufene Verbreiterung der aus Einzelmolekülzeitreihen berechneten Verteilungen von Bindungskonstanten $K_A$ mathematisch beschrieben. Eine heterogene Verteilung von Geschwindigkeitskonstanten lässt auf eine zerfurchte Energielandschaft schließen, in der mehrere d3'EBS1*/IBS1*-Konformationen gleichzeitig vorkommen. Die Zugabe von $M^{2+}$ begünstigt das Auftreten von solcher Substrukturen, die sich durch mikroskopische Unterschiede in der Faltung unterscheiden müssen.

Um das Phänomen kinetischer Heterogenität besser zu verstehen, wurden die $Mg^{2+}$-Bindungstaschen in den kürzlich veröffentlichten Strukturen von d3'EBS1* und d3'EBS1*/IBS1* mit Hilfe von NMR genauer untersucht (Abbildung 11.3A). Titrationen mit $[Co(NH_3)_6]^{3+}$

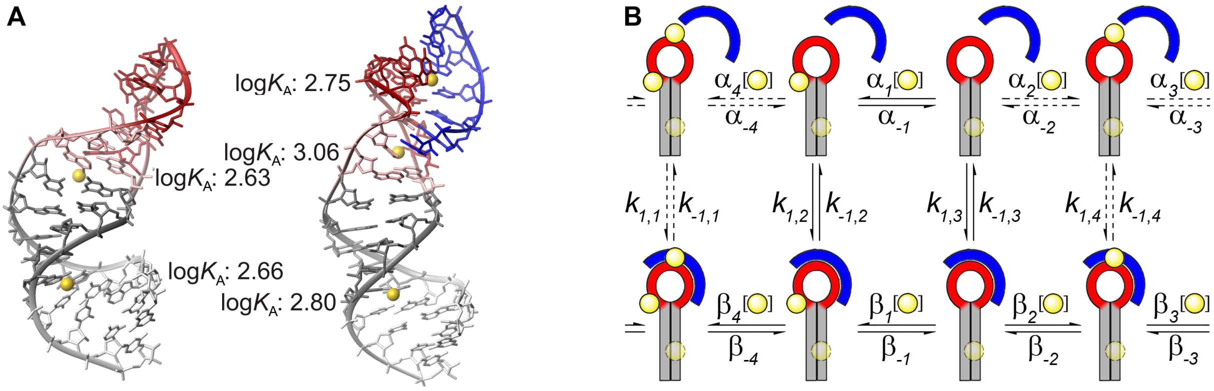


Figure 11.3: (A) Struktur und Metallionenbindungsstellen entlang der d3'EBS1*- und d3'EBS1*/IBS1*-Stammschleifenstruktur. Abgebildet sind jeweils die energetisch günstigsten von insgesamt 20 berechneten Strukturen. Die gelben Kugeln repräsentieren experimentell bestimmte $Mg^{2+}$-Bindungsstellen, für die jeweils eine Abschätzung der Bindungskonstante angegeben ist. (B) Modellvorschlag zur Erklärung der durch Kationen hervorgerufenen Heterogenität. Physiologische $Mg^{2+}$- oder $Ca^{2+}$-Konzentrationen führen zum gleichzeitigen Auftreten verschiedener d3'EBS1*-Konformere, die sich hinsichtlich der Besetzung der Kationenbindungsstellen und der IBS1*-Anbindungs- und Dissoziationsraten unterscheiden ($k_1$, $k_{-1}$). Die Umwandlung der Konformere erfolgt in Abhängigkeit der Kationen-Austauschraten ($\alpha$, $\beta$). In diesem Modell gilt die Annahme, dass der Einfluss der Besetzung der in der Nähe des G4-C26-Basenpaars befindlichen Kationbindungsstelle auf die Wechselwirkung der beiden RNS-Sequenzen vernachlässigbar ist.

und $Mn^{2+}$ deuten darauf hin, dass die beiden direkt an der EBS1*/IBS1*-Wechselwirkung beteiligten Metallionen teilweise dehydratisiert sind. Direkte Koordination von Metallionen an RNS führt im Allgemeinen zu stabilerer Anbindung und einer Verringerung der Austauschraten. Die Ergebnisse der Abschätzung der Affinität der Metallionen zu den jeweiligen Bindungsstellen deuten zudem darauf hin, dass diese im physiologischen $M^{2+}$-Konzentrationsbereich im Mittel nur unvollständig besetzt sind (Abbildung 11.3A).

Diese neuartige Kombination von smFRET und NMR legt deutlich dar, dass intermolekulare Heterogenität mit unvollständiger Besetzung von Metallionenbindungsstellen einhergeht. Seit der Entwicklung der Einzelmolekülspektroskopie wird regelmäßig von Strukturänderungen in RNS-Molekülen berichtet, die sich durch multiexponentielle Geschwindigkeitsgesetze auszeichnen. Es herrscht allerdings bis heute Uneinigkeit bezüglich der Erklärung dieses Phänomens (vgl. Kapitel 6). Unsere experimentellen Daten deuten darauf hin, dass mikroskopische Unterschiede zwischen einzelnen RNS-Konformeren höchstwarscheinlich von Unterschieden hinsichtlich der Besetzung der Metallionenbindungsstellen herrühren (Abbildung 11.3B). Somit liefert diese Studie eine Erklärung für eine seit langem bestehende Fragestellung in der Biophysik.

### Systematische Untersuchung des Einflusses zweiwertiger Metallionen auf die Ausbildung einer RNS-Tertiärstruktur mittels Einzelmolekülspektroskopie

Wechselwirkungen zwischen Metallionen und RNS findet an den negativ geladenen Phosphatgruppen entlang des Rückrads und an den exo- und endozyklischen Stickstoff- und Sauerstoffatomen der Nukleobasen statt (Abbildung 11.4A). Die Affinität verschiedener divalenter Metallionen für diese Bindungsstellen wurde mit Hilfe kleiner organischer Moleküle untersucht (Abbildung 11.4B). Dabei folgt die Stabilität von Acetat-, Imidazol- und Benzimidazolkomplexen der Reihe $Ba^{2+} < Sr^{2+} < Ca^{2+} < Mg^{2+} < Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} << Cu^{2+} > Zn^{2+} < Cd^{2+}$. Dieser Stabilitätstrend wurde bereits zuvor mit anderen Liganden beobachtet und ist unter dem Namen Irving-Williams-Reihe bekannt. Die Stabilität von Phosphat-Monoesterkomplexe erreicht im Gegensatz dazu in Anwesenheit von $Mn^{2+}$ ein lokales Maximum (Abbildung 11.4B). Dieses Kapitel beschreibt die erste systematische Untersuchung des Einflusses zweiwertiger Metallionen der erweiterten Irving-Williams-Reihe auf die Ausbildung einer RNS-Tertiärstruktur mittels smFRET.

Einzelmolekül-FRET-Messungen wurden in 100 mM $K^+$ und 1 mM $M^{2+}$ (M = Ba, Sr, Ca, Mg, Mn, Co, Ni oder Cd) durchgeführt. Kumulierte *FRET*-Histogramme und die Analyse temperaturabhängiger UV-Schmelzkurven zeigen, dass sich die verschiedenen Metallionen sehr unterschiedlich auf die Stabilität des d3'EBS1*/IBS1*-Komplexes auswirken. Entlang der erweiterten Irving-Williams-Reihe nimmt die Stabilität des d3'EBS1*/IBS1*-Komplexes zu $Mn^{2+}$ hin zu, gefolgt von einem Abfall in Anwesenheit von $Co^{2+}$ und $Ni^{2+}$(Abbildung 11.5A). UV-Schmelzkurven deuten ferner darauf hin, dass die Stabilität des RNS-RNS Komplexes in Gegenwart von $Zn^{2+}$ und $Cd^{2+}$ sehr hohe Werte erreicht. Unter diesen Bedingungen aufgenommene Einzelmolekülzeitreihen sind wegen der schweren Löslichkeit der Kationen, unzureichender Passivierung der Oberfläche und Wechselwirkung der Metallionen mit BSA nicht auswertbar und/oder mit Vorsicht zu interpretieren.

Um den Einfluss der einzelnen Metallionen auf die Wechselwirkung der beiden Sequenzen genauer zu untersuchen, wurde eine kinetische Analyse der Einzelmolekülzeitreihen durchgeführt. Aufgrund der in Kapitel 5 und 6 beschriebenen, metallionenabhängingen Heterogenität wurden die histogrammierten Verweilzeiten mittels gestreckter Exponentialfunktionen angepasst, die statt diskreter Geschwindigkeitskonstanten eine Verteilung von Raten annehmen. Die relativen Anbindungsgeschwindigkeiten nehmen entlang der erweiterten Irving-Williams-Reihe stetig zu und erreichen ihren höchsten Wert in Anwesenheit von $Ni^{2+}$ (Abbildung 11.5B). Im Gegensatz dazu erreichen die relativen Dissoziationsgeschwindigkeiten in Gegenwart von $Mn^{2+}$ ein Minimum, gefolgt von einer Zunahme hin zu $Co^{2+}$ und $Ni^{2+}$ (Abbildung 11.5B). Die Ergebnisse zeigen demnach deutlich, dass der Cofaktor sowohl die Anbindungs- als auch die Dissoziationsreaktion beeinflusst. Hierbei muss allerdings hervorgehoben werden, dass sich die Geschwindigkeit der Anbindung für die hier untersuchten Metallionen maximal um einen Faktor von 2 ändert, wohingegen die Dissoziationsraten eine komplette Größenordnung umfassen. Die ausgeprägten Unterschiede hinsichtlich der Stabilität des d3'EBS1*/IBS1*-Komplexes sind folglich hauptsächlich auf eine metallionenabhängige Stabiliserung und weniger auf eine Beschleunigung der Anbindungsreaktion zurückzuführen.

Der Vergleich der experimentellen Daten mit dem Verlauf der Komplexstabilitäten der hier untersuchten Metallionen macht deutlich, dass die Stabilität des d3'EBS1*/IBS1*-Komplexes von der Affinität des Cofaktors zum Phosphatliganden abhängt (Abbildungen 11.4 und 11.5).

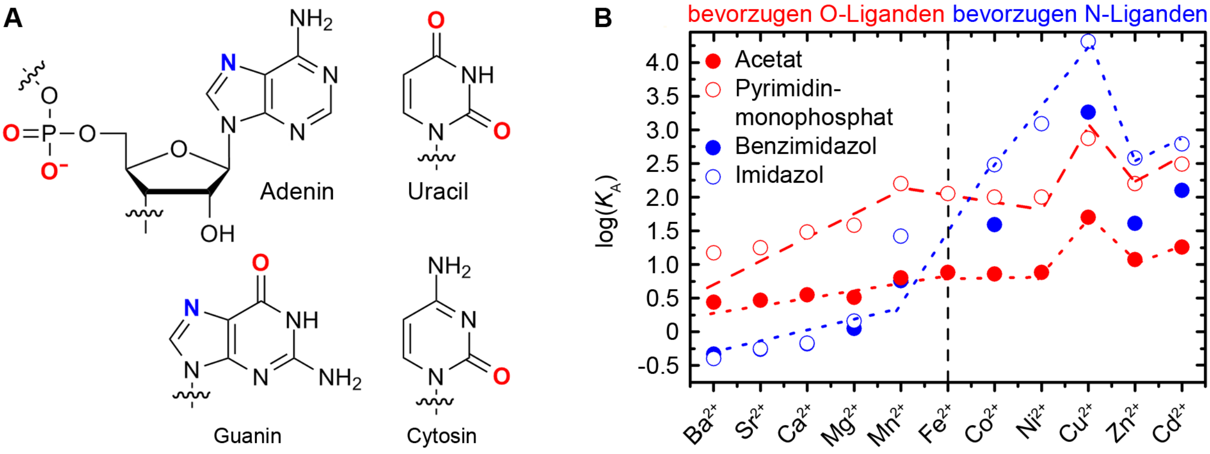


Figure 11.4: Wechselwirkung von Metallionen und RNS. (A) Übersicht über die wichtigsten in RNS vorkommenden Kationbindungsstellen (farbig hervorgehobene Atome). (B) Untersuchung der Stabilität spezifischer Metallionen-RNS-Wechselwirkungen unter Zuhilfenahme kleiner organischer Verbindungen. Während die relativen Stabilitäten von Acetat-, Imidazol und Benzimidazolkomplexen mit der Irving-Williams-Reihe übereinstimmen, weisen Phosphat-Monoester-Komplexe in Anwesenheit von $Mn^{2+}$ ein lokales Stabilitätsmaximum auf. Die Linien dienen als optische Orientierungshilfe.

Es ist demnach bergründet anzunehmen, dass einer stabilen Bindung zwischen d3'EBS1* und IBS1* die spezifische Koordination des zweiwertigen Metallions an Phosphatgruppen zugrunde liegt. Im Gegensatz dazu folgen die Anbindungsraten der klassischen Irving-Williams-Reihe. Dies impliziert, dass die Affinität des Metallions zu den Phosphatgruppen keinen nennenswerten Einfluss auf die Anbindung von IBS1* ausübt. Wahrscheinlich hängt die Anbindungsgeschwindigkeit stattdessen von der Fähigkeit des Metallions ab, negative Ladungen unspezifisch abzuschirmen. Zusammenfassend validiert diese Studie smFRET als Methode zur Untersuchung des Einflusses von Kationen auf die Faltung von RNS. Es wird zudem gezeigt, dass die Bildung eines RNS-RNS-Duplexes direkt von den Koordinationseigenschaften der in Lösung befindlichen Metallionen abhängt.

**Randomisierung in der Analyse von Einzelmolekül-FRET-Daten**

Da es Einzelmolülspektroskopiemethoden wie smFRET ermöglichen, die Beobachtung der Faltung von Biomolekülen in Echtzeit zu verfolgen, erfreuen sie sich in den letzten Jahren steigender Beliebtheit. Begrenzte Beobachtungsdauer und das Auftreten heterogener Kinetiken und Konformationen führen allerdings häufig zu ausgeprägten intermolekularen Unterschieden bezüglich der relativen Häufigkeit der *FRET*-Zustände und den dazugehörigen Umwandlungsraten (vgl. Kapitel 5 und 6). Die Quantifizierung solcher Unterschiede ist Vorausssetzung für die Durchführung statistischer Tests. Bisher wurden allerdings sowohl die Variabilität als auch statistische Tests in Einzelmolekülstudien auf Kosten robuster biologischer Interpretation der

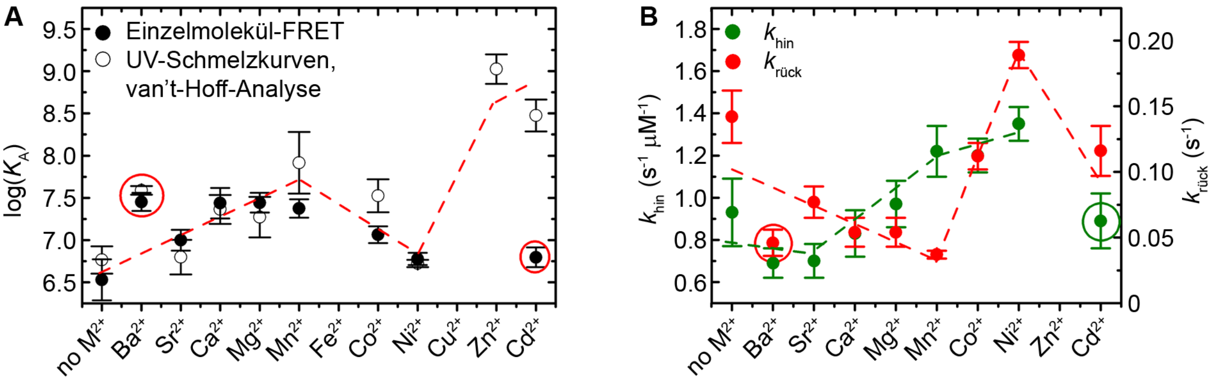


Figure 11.5: Thermodynamische und kinetische Kenngrößen der Wechselwirkung zwischen d3'EBS1* und IBS1* in Abhängigkeit des Metallionen-Cofaktors. (A) Bestimmung der Stabilitätskonstanten des d3'EBS1*/IBS1*-Komplexes in Anwesenheit von 100 mM $K^+$ und 1 mM $M^{2+}$ mittels Einzelmolekül-FRET und Analyse von UV-Schmelzkurven. (B) Anbindungs- und Dissoziationsraten von IBS1* in Gegenwart verschiedener zweiwertiger Metallionen. Die Geschwindigkeitskonstanten wurden durch Anpassen getreckter Exponentialfunktionen an Verweilzeit-Histogramme erhalten. Die Streuung in Einzelmolekülexperimenten wurde durch Randomisierung abgeschätzt ($3\sigma_{\mathrm{bootstrap}}$). Der Fehler in UV-Schmelzkurven wurde mittels biologischer Replikate ermittelt ($1\sigma$). Die Linien dienen als optische Orientierungshilfe.

experimentellen Daten weitgehend außer Acht gelassen. Dieses Kapitel liefert das Rüstzeug zur Abschätzungung intermolekularer Unterschiede mittels Randomisierung und der Durchführung von statistischen Tests in smFRET-Experimenten.

Bootstrapping (*engl.* “Stiefelschlaufe”) ist eine Methode der Randomisierung, in der wiederholt die Probenstatistik auf der Grundlage einer einzigen Stichprobe von *n* Replikaten berechnet wird. Bootstrapping liefert dadurch eine Abschätzung des Mittelwerts und der dazugehörigen Streuung. In der Einleitung wird zunächst mathematisch gezeigt, dass Bootstrapping mit routinemäßig in der Untersuchung von Einzelmolekülzeitreihen eingesetzen Analysemethoden kombinierbar ist (Abbildung 11.6). Anschließend wird mit Hilfe simulierter Datensätze die Robustheit des Algorithmus in Kombination mit verschiedenen Methoden zur thermodynamischen und kinetischen Analyse von Einzelmolekülzeitspuren quantitativ getestet. Die Ergebnisse zeigen, dass sich die mittels eines Bootstrapping abgeschätzten Mittel- und Streuungswerte systematisch mit den theoretischen Sollwerte decken. Voraussetzung ist hi-

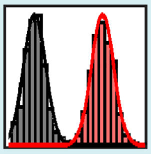
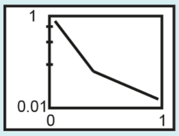
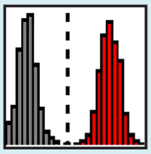
| | **Thermodynamische Analyse** | **Kinetische Analyse** |
|---|---|---|
| **Methode 1** | **Randomisierung und Regression** normierter cumulierter *FRET*-Histogramme $\alpha(A,b,\sigma)$.<br>**Eingabe**: 1D-FRET-Histogramme, erstellt aus Einzelmolekülzeitreihen.<br>**Ausgabe**: Mittelpunkt, Breite, Amplitude und mittlere Häufigkeit der einzelenen FRET-Verteilungen inkl. der Variabilität $\sigma_{\text{bootstrap}}$.<br> | **Randomisierung und Regression** normierter cumulierter Verweilzeit-Histogramme $\alpha(A,b,\sigma)$.<br>**Eingabe**: Verweilzeiten aus Einzelmolekülzeitreihen, bestimmt durch Schwellwertanalyse oder HMM.<br>**Ausgabe**: Zerfallskonstanten, optional: Amplitude und Streckungsfaktor inkl. der Variabilität $\sigma_{\text{bootstrap}}$.<br> |
| **Methode 2** | **Randomisierung und Mittelung** von Verweilzeiten in einem best. FRET-Zustand.<br>**Eingabe**: Verweilzeiten aus Einzelmolekülzeitreihen, bestimmt durch Schwellwertanalyse oder HMM.<br>**Ausgabe**: Mittlere Verweilzeit in den einzelenen FRET-Zuständen, mittlere Häufigkeit, Variabilität $\sigma_{\text{bootstrap}}$.<br> | - |
| **Methode 3** | **Randomisierung und Integration** von FRET-Verteilungen in normierten kumulierten FRET-Histogrammen<br>**Eingabe**: Aus Einzelmolekülzeitreihen erstellte Histogramme.<br>**Ausgabe**: Mittlere Häufigkeit, Variabilität $\sigma_{\text{bootstrap}}$.<br> | - |

Figure 11.6: Übersicht der verschiedenen in BOBA FRET implementierten Funktionen. Ebenfalls angegeben sind die entsprechenden Ein- und Ausgabenvariablen. Für eine genau mathematische Beschreibung der angegebenen Routinen sei auf Kapitel 8 verwiesen.

erbei allerdings, dass die randomisierten Eingabe-Werte durch eine der jeweiligen Datenlage angepassten Methode bestimmt werden. Abschließend die Methode auf einen Satz experimenteller Daten angewendet und somit gezeigt, dass divalente Metallionen einen statistisch signifikanten Einfluss auf die mit der Wechselwirkung von d3'EBS1* und IBS1* verbundenen thermodynami- schen und kinetischen Kenngrößen haben.

Zusammenfassend validieren die hier durchgeführten mathematischen Ausführungen, Simulationen und Experimente Bootstrapping zur Quantifizierung intermolekularer Unterschiede in smFRET-Zeitreihen. Ein leicht zu implementierender Algorithmus mit grafischer Benutzeroberfläche zur Randomisierung von smFRET Daten wird zudem zum kostenlosen Download angeboten. Da die Abschätzung intermolekularer Unterschiede den Weg für robuste biologische Interpretation von smFRET-Experimenten frei macht, gehen wir davon aus, dass diese Methode breite Anwendung finden wird.

# Chapter 12

# Appendix

## 12.1 Instrumentation for TIR smFRET imaging

Figure 12.1 depicts the custom TIRF microscope used for smFRET experiments presented in this work. The optical components of the setup were optimised for using the Cy3/Cy5 FRET pair (Figure 12.2), though other fluorophores with similar spectral properties could in principle be employed as well, for example tetramethylrhodamine and Alexa Fluor ® 647.

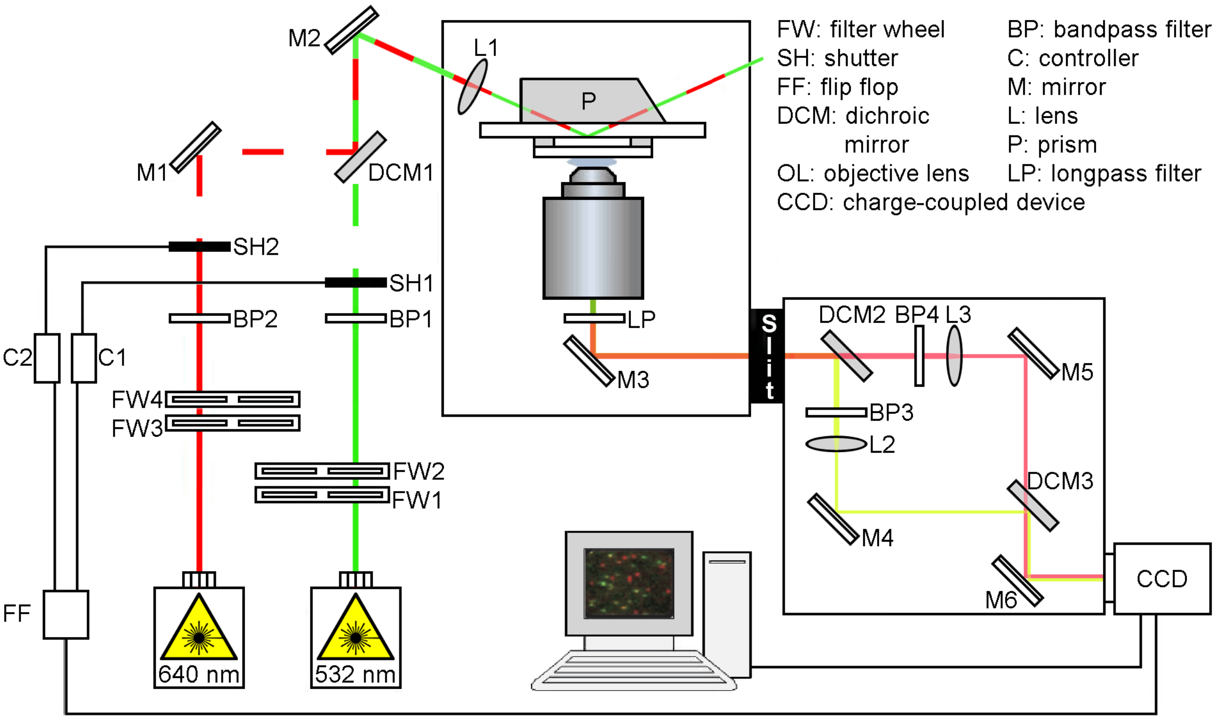


Figure 12.1: Scheme of the TIRF microscope used for single-molecule imaging.

### 12.1.1 Excitation

Two continuous-wave lasers (CL532-050-L and CRL-DL640-035-US-0.25, low noise, CrystaLaser lc., Reno, NV, USA) were attenuated in intensity using filter wheels mounted with neutral density filters (FW: Laser2000 GmbH, Weling, Germany) and cleaned up using bandpass filters (BP1, BP2: AHF Analysentechnik AG, Tübingen, Germany, Figure 12.2). A pair of synchronised mechanical shutters (SH, C: LS-3, Uniblitz, Rochester, NY, USA) and a self-built flip-flop were used for alternating laser excitation (ALEX) of fluorophores as described in [125]. This technique permits to monitor FRET efficiencies and fluorophore stoichiometry in real-time, and allows to detect photophysical artifacts such as doubly-labelled molecules, blinking, and spectral shifts [146, 354, 393]. A system consisting of dielectric mirrors (M1, M2: Laser2000 GmbH, Weling, Germany) and a dichroic beam splitter (DCM1: AHF Analysentechnik, Tübingen, Germany) was used for aligning the optical paths of the lasers, followed by focussing the incident beam onto the prism (P: Melles Griot GmbH, Bensheim, Germany) and the sample chamber using a biconvex lens (L1: Newport Spectra Physics GmbH, Germany). Sample chambers were built from quartz slides (G. Finkenbeiner Inc., Waltham, MA, USA), double-sides sticky tape (3M, Cergy Pontoise, France), and glass coverslips (Roth AG, Karlsruhe, Germany), followed by sealing with epoxy glue as described [170].

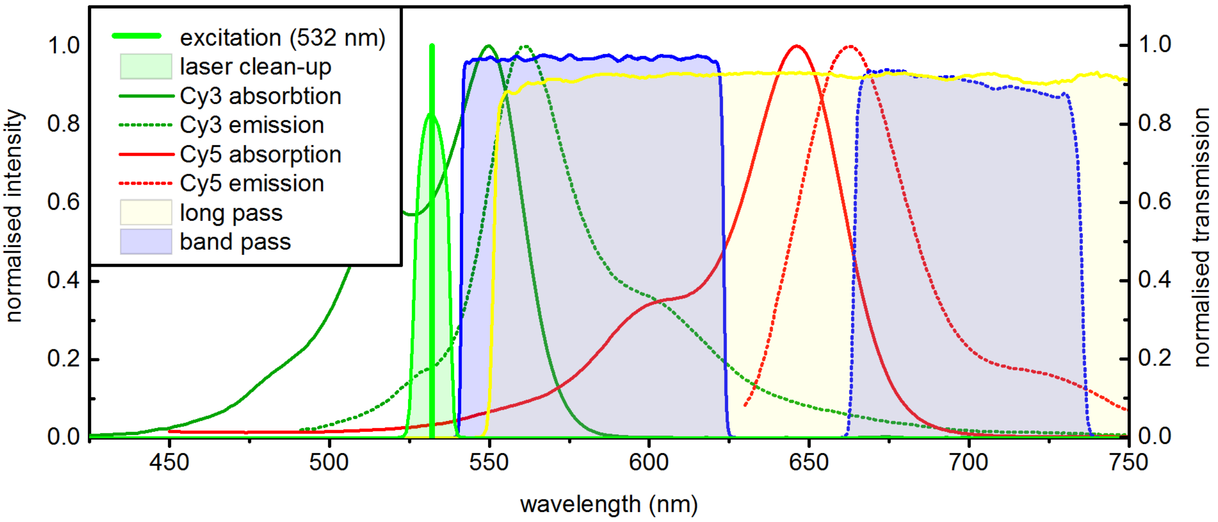


Figure 12.2: Fluorochrome spectra and selected filters used in the TIRF setup used shown in Figure 12.1. Cy3 and Cy5 absorption and emission spectra correspond to experimental data recorded by Dr. Danny Kowerko. Transmission graphs for the laser clean-up filter (BP1), the long pass filter (LP), and the band pass filters (BP3, BP4) were provided by AHF Analysentechnik AG, Tübingen, Germany. Not shown: The laser clean-up filter for excitation at 640 nm (BP2), dichroic mirrors (DCM1, DCM2, DCM3).

### 12.1.2 Detection

The prism-based TIRFM inverted microscope was purchased from Olympus Schweiz AG (Olympus IX71, Volketswil, Switzerland) and equipped with a long-pass filter (Lot-Oriel AG, Romanel-sur-Morge, Switzerland) to reject scattered laser light (Figure 12.2). Fluorophore emission was spectrally separated using a sequential alignment of dichroic beam splitters (DCM2, DCM3: Lot-Oriel AG, Romanel-sur-Morge, Switzerland) and dielectric mirrors (M3, M4, M5: Laser2000 GmbH, Weling, Germany). Bandpass filters (BP3, BP4: AHF Analysentechnik AG, Tübingen, Germany) were used attenuate leakage of Cy3 emission into the Cy5 detection channel and to suppress inelastic Raman scattering from water (Figure 12.2) [161, 474]. Finally, Cy3 and Cy5 emission levels were projected side-by-side onto an EM-CCD detector using biconvex lenses (Newport Spectra Physics GmbH, Darmstadt, Germany). All components were mounted onto an optical top equipped with vibrational isolation (CleanTop II 1200 x 3600 x 300 mm, General Microtechnology and Photonics S.A., Renens, Switzerland).

## 12.2 Cy3 and Cy5 chemical structure and photophysics

Cyanine generally consist of two five-membered heterocycles, an uncharged and a positively charged one, each of which is fused to a benzene ring (*nitrogen centre*) [475]. Both nitrogen centres are linked by a conjugated chain of an odd number of carbons, that are generally in all-*trans* conformation, though, they may occasionally adopt the *cis*-form [475]. The cyanine dyes used for all fluorescence experiments presented in this work are Cy3 and Cy5, in which

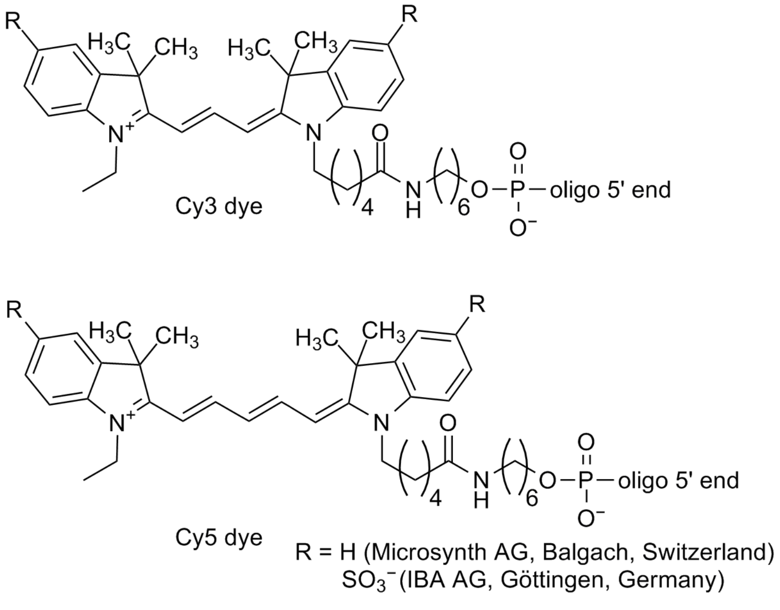


Figure 12.3: Dyes used in single-molecule and bulk fluorescence experiments presented herein. Structures communicated by Marc Niwar, IBA GmbH, Göttingen, Germany.

the nitrogen centres are spaced by 3 or 5 carbons, respectively (Figure 12.3). Alkyl groups are attached to the nitrogen centres to decrease the reactivity of the dyes, which in turn enhances their photostability [139]. Both fluorophores are conjugated to the 5' end of the RNA via a $(CH_2)_{10}$ chain interrupted by an amide bond, which is formed during the process of synthesis upon reaction of an activated cyanine dye N-hydroxysuccinimidyl ester and a primary amine. Importantly, IBA AG (Göttingen, Germany) uses sulfonated dyes with a negative net charge and enhanced solubility in water, while labelled RNAs from Microsynth AG (Balgach, Switzerland) are conjugated to unsulfonated fluorophores. As depicted in Figure 12.2, Cy3 is fluorescent in the green range of the visible spectrum ($\lambda_{\text{exc}} \sim 550$ nm, $\lambda_{\text{em}} \sim 570$ nm), while Cy5 fluoresces in the red region ($\lambda_{\text{exc}} \sim 650$ nm, $\lambda_{\text{em}} \sim 670$ nm).

### 12.2.1 Dynamic quenching of Cy3 and Cy5 fluorescence induced by $M^{2+}$

Fluorescence quenching denotes any process that decreases the fluorescence intensity and it can be caused by a variety of molecular interactions, including excited-state reactions, molecular rearrangements, energy transfer, ground-state complex formation, and collisional quenching (see also Section 3.2.3) [109]. In the context of this study, a decrease in Cy3 quantum yield would be accompanied by a decrease of $R_0$, and hence an artefactual decrease of *FRET* [134]. In the case of Cy5 quenching, *FRET* values in single molecule experiments would also change. Cations, in particular transition metal ions, are known to quench fluorescence through diverse

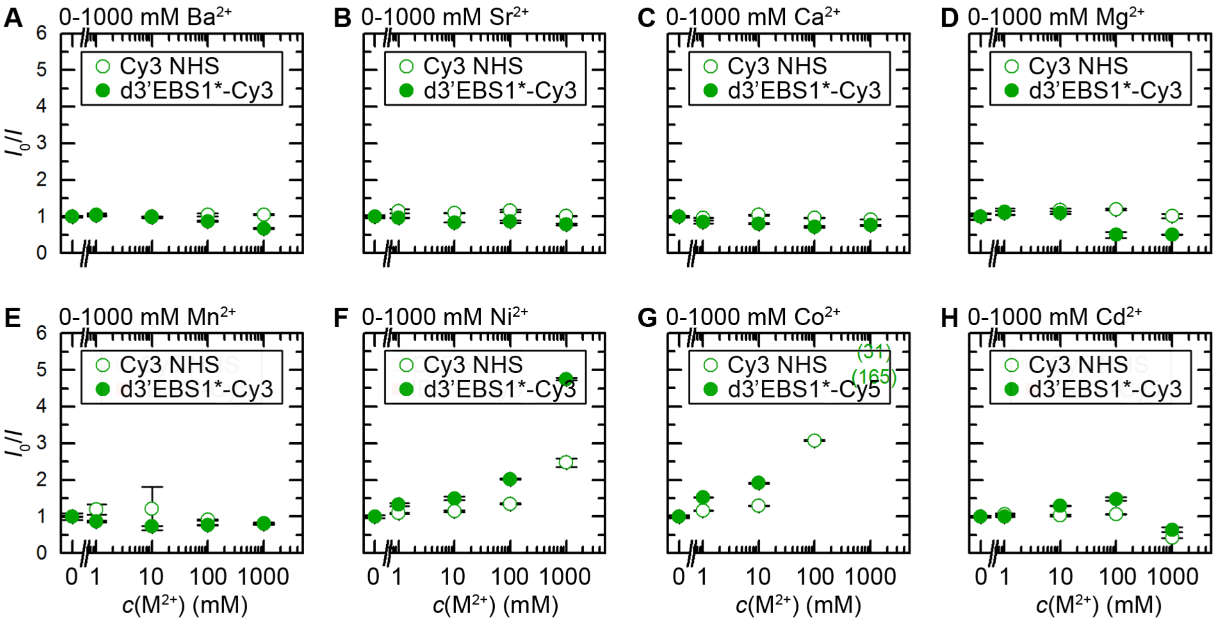


Figure 12.4: Stern-Vollmer plots for quenching of Cy3 fluorescence by different divalent metal ions as determined by fixed wavelength scans at $\lambda_{\text{exc}} = 530$ nm. Imaging buffer: 50 mM MOPS, 100 mM $KNO_3$, 1 mM Trolox, 1x OSS, pH 6.9 and divalent metal ions as indicated. Error bars correspond to the standard deviation $1\sigma$.

mechanisms, the investigation of which would go beyond the scope of this study [115, 476, 477]. This section provides an account of the concentrations at which the earth alkali metal ions $Ba^{2+}$, $Sr^{2+}$, $Ca^{2+}$, and $Mg^{2+}$, as well as the transition metal ions $Mn^{2+}$, $Ni^{2+}$, $Co^{2+}$, and $Cd^{2+}$ quench of Cy3 or Cy5 fluorescence to an extent that may potentially bias FRET studies. Furthermore, the effect of dye conjugation was investigated by comparing the fluorescence intensity of cyanine dyes in their commercially available NHS ester form (*Cy3 NHS*, *Cy5 NHS*) and upon attachment to RNA. It should be noted, however, that NHS esters rapidly hydrolyse in aquaeous solution forming cyanine dyes with a carboxyl moiety [478].

Figures 12.4 and 12.5 show Stern-Vollmer plots of Cy3 and Cy5 fluorescence, where $I_0$ and $I$ are the absolute intensities in the absence and presence of $M^{2+}$. Fluorescence intensities of NHS esters and RNA-attached dyes are generally in excellent agreement, demonstrating that bioconjugation has a negligible effect on fluorophore quantum yields. However, at increasing cation concentrations of $Sr^{2+}$, $Ca^{2+}$, $Mg^{2+}$ and $Cd^{2+}$, minor differences are observed regarding the quantum yields of Cy3 NHS and EBS1*-Cy3. Strong fluorescence quenching is observed in the presence of $Ni^{2+}$ and $Co^{2+}$. While $Co^{2+}$ displays a more pronounced effect on Cy3 emission than $Ni^{2+}$, both cations have a comparable effect on fluorescence of Cy5. In contrast, slight fluorescence enhancement is observed for Cy3 in the presence of $Mg^{2+}$ or $Cd^{2+}$. However, it should be noted that all of the aforementioned effects become significant solely at 100mM $M^{2+}$ or higher concentrations. $Co^{2+}$- and $Ni^{2+}$-induced quenching has been reported earlier using other fluorescent compounds (NBD, PPO, and perylenes) [118, 479, 480, 481, 482, 483]. Be-

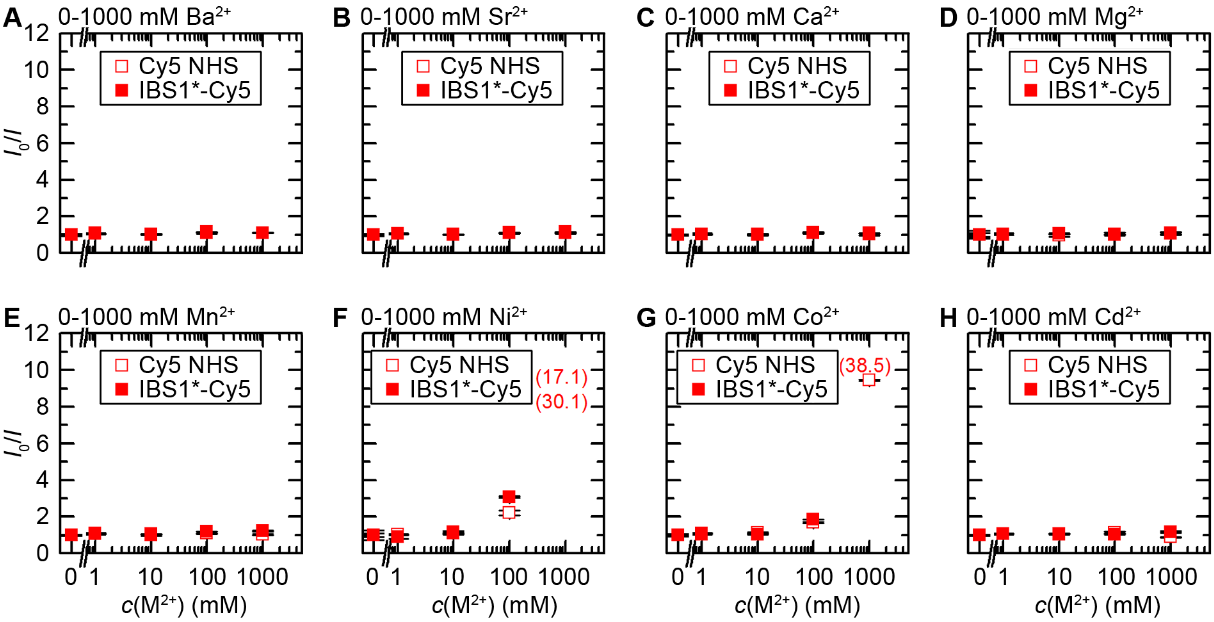


Figure 12.5: Stern-Vollmer plots for quenching of Cy5 fluorescence by different divalent metal ions as determined by fixed wavelength scans at $\lambda_{exc}$ = 600 nm. Imaging buffer: 50 mM MOPS, 100 mM $KNO_3$, 1 mM Trolox, 1x OSS, pH 6.9 and divalent metal ions as indicated. Error bars correspond to the standard deviation $1\sigma$.

sides collisional quenching, $Ni^{2+}$ has been demonstrated to quench perylene fluorescence by short range Dexter interaction, while $Co^{2+}$ quenches via long range FRET [118, 482]. Taken together, these data demonstrate that (i) conjugation of Cy3 and Cy5 to the RNA sequences used herein does not entail significant effect on fluorophore quantum yields and that (ii) fluorescence quenching by $M^{2+}$ is a negligible effect for the cations studied here when they are present at low millimolar concentrations. Hence, smFRET experiments are not expected to be convoluted with $M^{2+}$-induced Cy3/Cy5 fluorescence quenching at the $M^{2+}$ concentrations used in the context of this work.

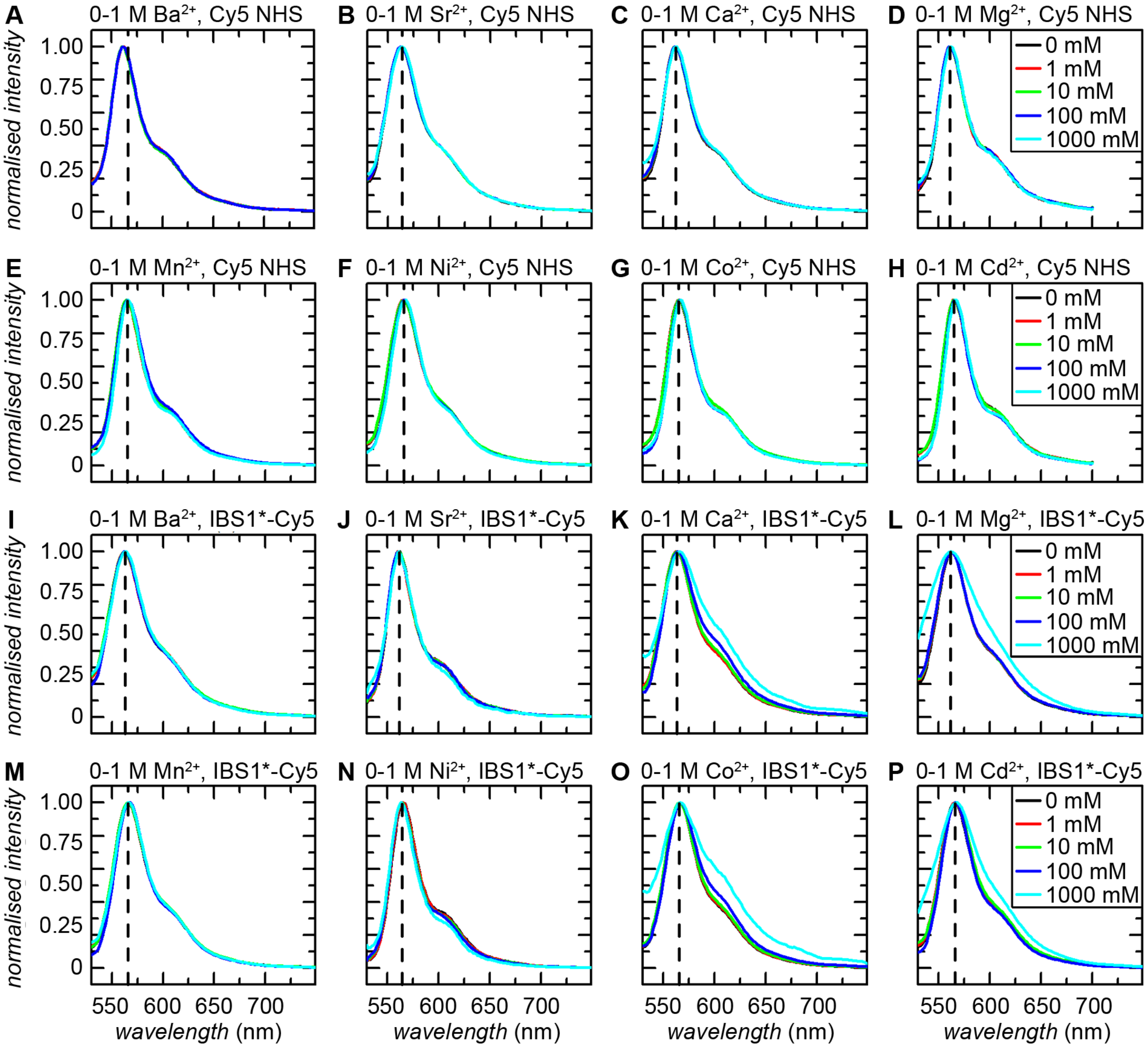


Figure 12.6: Normalised fluorescence spectra of free Cy3-NHS ester (A-H) and Cy3 conjugated to d3'EBS1* (I-P) at different $M^{2+}$ concentrations as determined upon excitation at $\lambda_{exc} = 510$ nm. Dashed lines indicate the emission maximum. Imaging buffer: 50 mM MOPS, 100 mM $KNO_3$, 1 mM Trolox, 1x OSS, pH 6.9 and divalent metal ions as indicated.

### 12.2.2 Shape of Cy3 and Cy5 fluorescence spectra in the presence of $M^{2+}$

The rate of energy transfer also depends upon the overlap integral of the emission spectrum of the FRET donor and the absorption spectrum of the acceptor [109]. Spectral shifts would consequently lead to fluctuations in the apparent *FRET* efficiency, an phenomenon that has been reported for ATTO647N and Alexa488 [146]. It is therefore important to provide evidence that cation-induced spectral shifts and spectral alterations are negligible. This section provides an account of the influence of $Ba^{2+}$, $Sr^{2+}$, $Ca^{2+}$, $Mg^{2+}$, $Mn^{2+}$, $Ni^{2+}$, $Co^{2+}$, and $Cd^{2+}$ on the shape of Cy and Cy5 fluorescence spectra. Potential effects of dye conjugation to RNA were also characterised.

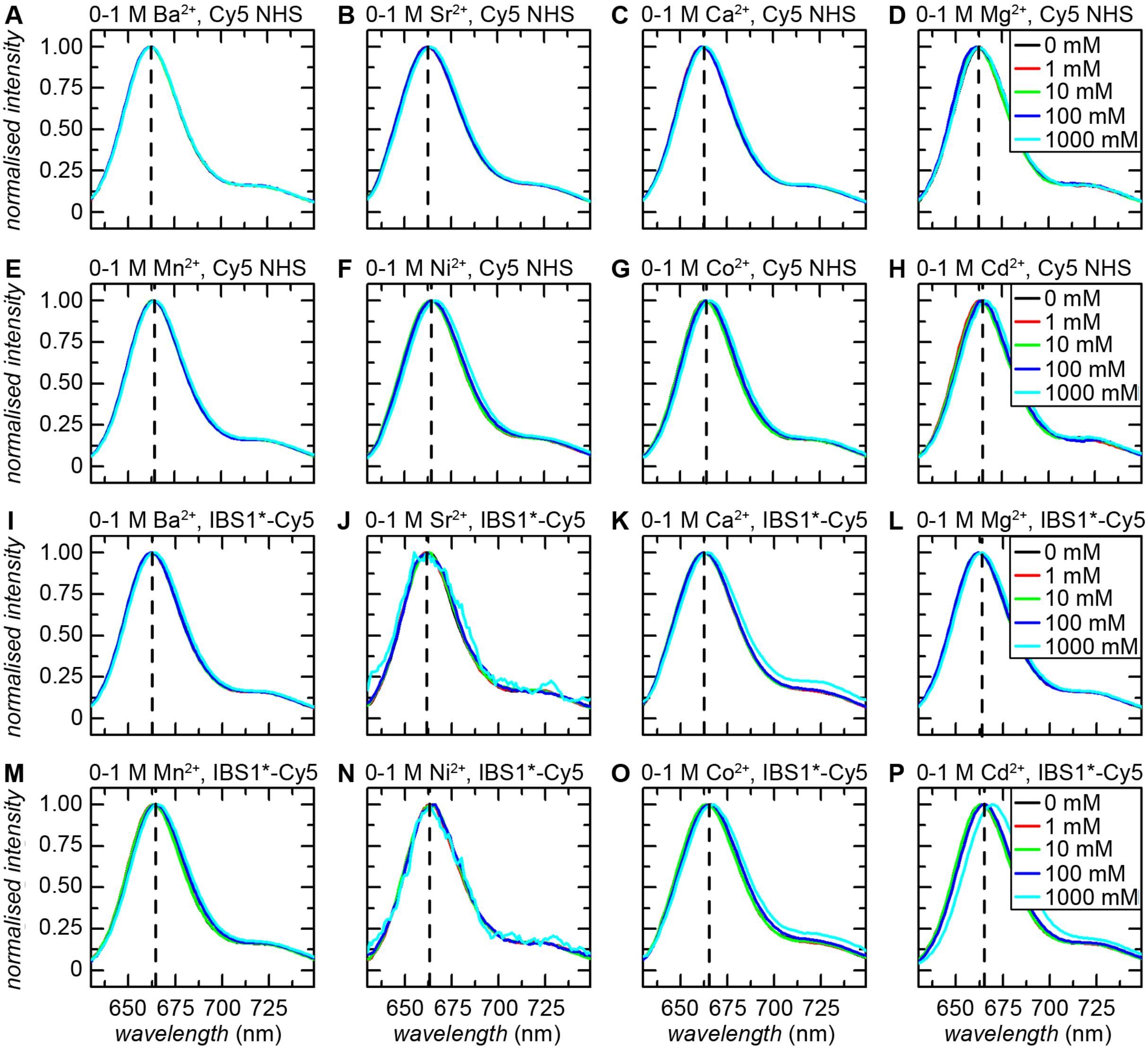


Figure 12.7: Normalised fluorescence spectra of free Cy5-NHS ester (A-H) and Cy5 conjugated to IBS1* (I-P) at different $M^{2+}$ concentrations as determined upon excitation at $\lambda_{\text{exc}}$ = 600 nm. Dashed lines indicate the emission maximum. Imaging buffer: 50 mM MOPS, 100 mM $KNO_3$, 1 mM Trolox, 1x OSS, pH 6.9 and divalent metal ions as indicated.

Normalised fluorescence spectra of Cy3 are shown in Figure 12.6. The spectral shape of NHS esters is largely unaffected by the presence of divalent metal ions at the range of concentrations studied herein. Minor broadening was observed at 1 M $Co^{2+}$ and $Cd^{2+}$, though differences in fluorescence intensity did not exceed 5 %. Emission spectra of Cy3-labelled RNA were in excellent agreement with the spectra measured for unconjugated fluorophores. As for Cy3 NHS, high amounts of $Co^{2+}$ and $Cd^{2+}$ induce spectral broadening, whereby differences reach values of up to 33 %. A similar effect is observed at high $Ca^{2+}$ and $Mg^{2+}$ concentrations, but to a lesser extent. Furthermore, a slight shift in the emission wavelength is observed upon Cy3 conjugation ($\Delta\lambda_{\text{em, max}} \leq 2$ nm). Figure 12.7 shows normalised Cy5 emission spectra recorded in the presence of various metal ion at different abundance. Emission profiles of NHS Cy5 are virtually unaffected by changes in the composition of the imaging buffer, both with regard to shape and fluorescence maxima. Increasing $Cd^{2+}$ concentration does, however, entail a slight batochromic shift ($\Delta\lambda_{\text{em, max}} \leq 3$ nm). Unlike for Cy3, linking Cy5 to its cognate RNA sequence does not lead to cation-dependent broadening of emission profiles. Instead, the red-shift was then consistently present and increasing along the Irving-Williams Series. It reaches a maximum at 1 M $Cd^{2+}$ ($\Delta\lambda_{\text{em, max}} = 7$ nm).

In summary, these results suggest that the electronic structure of Cy3 is affected by high concentrations of $Co^{2+}$ and $Cd^{2+}$. Even though $Ni^{2+}$ strongly quenches Cy3 fluorescence (see also: Section 12.2.1), its emission profile remains largely unaffected upon addition of $Ni^{2+}$, even at high concentrations. Consequently, the mechanisms by which these transition metal ions bring about fluorescence quenching are most likely dissimilar. Cation-dependent broaden-

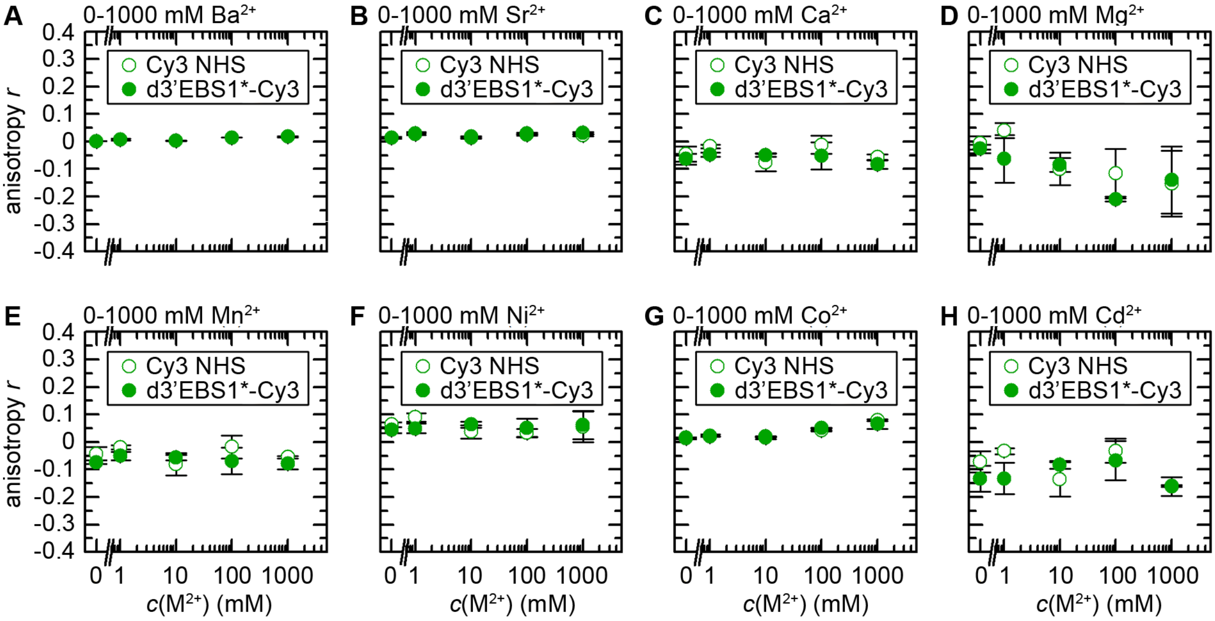


Figure 12.8: Metal ion-dependent anisotropy of Cy3 as determined by fixed wavelength scans at $\lambda_{\text{exc}}$ = 530 and $\lambda_{\text{em}}$ = 560/580 nm. Imaging buffer: 50 mM MOPS, 100 mM $KNO_3$, 1 mM Trolox, 1x OSS, pH 6.9 and divalent metal ions as indicated. Error bars correspond to the standard deviation $1\sigma$.

ing of fluorescence profiles may be caused by inhomogeneous broadening, *i.e.* non-equivalence of chromophore environement resulting in a distribution of solute-solvent interaction energies [484]. A thorough discussion of this phenomenon would go beyond the scope of this study. Most importantly, cation-chromophore interactions seem to be virtually absent at concentrations relevant for the smFRET experiments presented herein and/or their effect on fluorescence spectra is not significant.

### 12.2.3 Static Cy3 and Cy5 fluorescence anisotropy in the presence of $M^{2+}$

When a fluorophore is excited by linearly polarised light, the resulting emission is often also polarised, a phenomenon that is referred to as fluorescence anisotropy ($r$, see also: Equations 3.7 and 3.8 in Section 3.2.5) [109]. Förster theory accounts for the dependence of $R_0$ on the relative orientation of the donor and acceptor transition moments by introducing the orientation factor $\kappa^2$ ($0 \leq \kappa^2 \leq 4$, see also: Equation 3.6 in Section 3.2.4). If fluorophores undergo isotropic reorientation in a time much shorter than the rate of donor de-excitation, $\kappa^2$ is well approximated by 2/3 and the orientation is not expected to bias the interpretation of *FRET* traces [129]. This section surveys whether conjugation to RNA and/or the presence of the divalent metal ions $Ba^{2+}$, $Sr^{2+}$, $Ca^{2+}$, $Mg^{2+}$, $Mn^{2+}$, $Ni^{2+}$, $Co^{2+}$, and $Cd^{2+}$ significantly alter the rotational behavior of Cy3 and Cy5 via quantification of static fluorescence anisotropy.

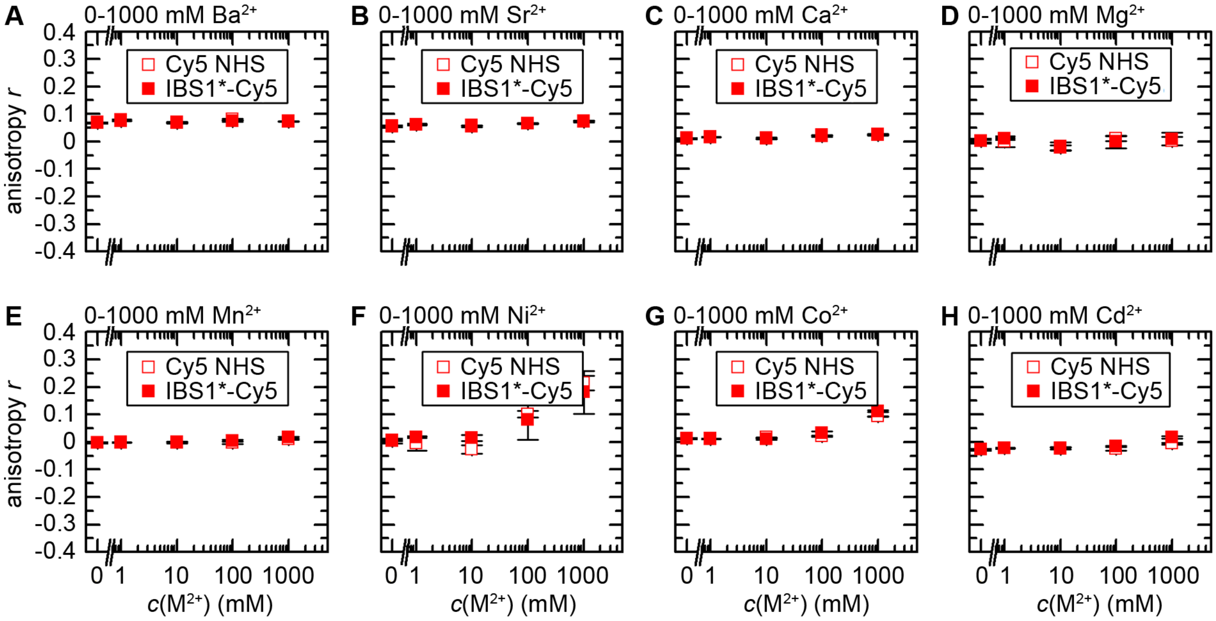


Figure 12.9: Metal ion-dependent anisotropy of Cy5 as determined by fixed wavelength scans at $\lambda_{exc}$ = 600 and $\lambda_{em}$ = 660/680 nm. Imaging buffer: 50 mM MOPS, 100 mM $KNO_3$, 1 mM Trolox, 1x OSS, pH 6.9 and divalent metal ions as indicated. Error bars correspond to the standard deviation $1\sigma$.

Figures 12.8 and 12.9 show Cy3 and Cy5 fluorescence anisotropy against metal ion concentration. Differences between anisotropies of free and RNA-tethered dyes are consistently within the accuracy of the method. Large cross-sample variabilities observed for Cy3 anisotropy in the presence of $Mg^{2+}$, $Mn^{2+}$, and $Co^{2+}$ are due to the fact that the instrument factor $G$ not only depends on selecting the well with the highest fluorescence intensity, but also on the automatic determination of the gain. Here, the gain value varies between individual measurements and adopted lower values in the presence of $Mg^{2+}$, $Mn^{2+}$, and $Co^{2+}$. Overall, G varied between 0.95 and 0.20, where decreasing gain values also led to lower *SNR*. Addition of $Ni^{2+}$ induces pronounced Cy5 fluorescence anisotropy, a phenomenon that has been applied in anisotropy-based $Ni^{2+}$ sensing using a different fluorophore [485].

Taken together, the divalent metal ions studied herein do not appear to have a significant impact on isotropic reorientation at low millimolar concentrations. High concentration of $Ni^{2+}$ does not only strongly quench both Cy3 and Cy5 emission, it also hinders free rotation of Cy5 attached to IBS1*.

### 12.2.4 The influence of $Mg^{2+}$ and $Ca^{2+}$ on Cy3 and Cy5 blinking and photobleaching as quantified by single-molecule fluorescence

Dyes commonly employed in single-molecule fluororescence studies often display fast photobleaching (irreversible inactivation after emission of roughly $10^5$ to $10^7$ photons), long-lived dark states (reversible inactivation, *blinking*), and intensity fluctuations [149, 166]. In time-binned single-molecule fluorescence experiments, these issues do not only limit the observation time and hence lead to considerable differences between single trajectories, they may also introduce uncertainty regarding the interpretation of the experimental data [166]. In particular, blinking is easily mistaken for a change in *FRET* and erroneously interpreted as a conformational change [486]. As photophysical issues are an intrinsic property of every fluorophore, control experiments aimed at their characterisation and quantification are crucial [146]. This section provides an overview of Cy3 and Cy5 photostability and blinking in the presence and the earth alkali metal ions $Ca^{2+}$ and $Mg^{2+}$ [111, 150]. For that purpose, fluorophores were conjugated to d3'EBS1* and immobilised on a quartz slide via a biotin-streptavidin linkage, followed by imaging single-molecule fluororescence with a TIRF microscope (10 frames/s, *observation time* = 400 s) at 50 mM MOPS (pH 6.90), 100 mM $K^+$, 1 % D-glucose, 1x OSS, 1 mM Trolox, and $M^{2+}$ at variable concentrations (see also: Section 12.1) [170].

Panels A and F in Figure 12.10 depict a set of representative Cy3 and Cy5 fluorescence time traces. Different scenarios are observed, including stable emission over the entire observation time of 400 s, photobleaching before the end of the movie, as well as brief blinking events. In contrast to Cy3 trajectories, multiple blinking states are clearly discernable in Cy5 time traces, *i.e* emission levels may drop to different levels. Figures 12.10B and G show histograms of the time that single fluorophores were imaged before photobleaching. It is evident that (i) there is no apparent statistical distribution and (ii) a large fraction of fluorophores do not bleach during the observation time. Figures 12.10D and I further illustrate that at all metal ion concentrations investigated herein, individual fluorophores strongly differ with regard to

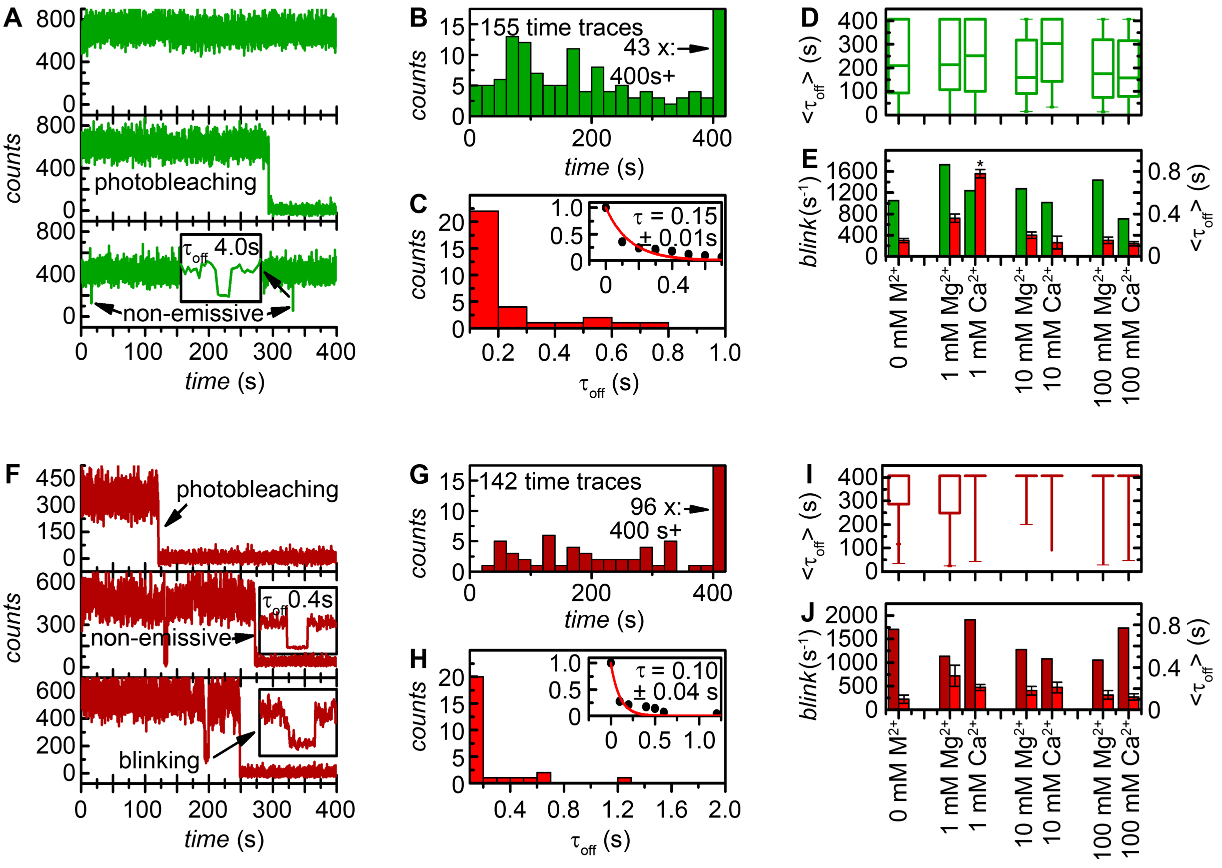


Figure 12.10: smFRET control experiments II. (A) Representative Cy3 emission time trajectories show stable emission (top), photobleaching (middle), and blinking (bottom). (B) Time until Cy3 photobleaching ($\tau_{\text{on}}$), representative data recorded at 0 mM $M^{2+}$. The histogram does not display a clearly defined statistical distribution. (C) Quantification of Cy3 blinking ($\tau_{\text{off}}$), representative data recorded at 0 mM $M^{2+}$. Red curve, inset: single-exponential decay. (D) Box plot of $M^{2+}$-dependent duration of stable Cy3 emission until photobleaching. Line: median; box: upper/lower quartile; whiskers: minimum/maximum. Dwell times cover almost the full range of the observation time with a median centered around 200 s. Divalent metal ions assessed herein do not have a significant influence on Cy3 photostability. (E) $M^{2+}$-dependent Cy3 blink frequency and average duration of the dark state. The average time spent in the dark state ($\langle\tau_{\text{off}}\rangle$) is close to the time resolution in all cases and does not appear to be influenced by $Mg^{2+}$ or $Ca^{2+}$. Blinking becomes more frequent at high $M^{2+}$ concentrations. (F) Representative Cy5 emission time trajectories unveil multistep blinking (middle, bottom). (G) Time until Cy5 photobleaching ($\tau_{on}$), representative data recorded at 0 mM $M^{2+}$. (H) Quantification of Cy5 blinking ($\tau_{\text{off}}$), representative data recorded at 0 mM $M^{2+}$. Red curve, inset: single-exponential decay. (I) Box plot of $M^{2+}$-dependent duration of stable Cy5 emission until photobleaching. (J) $M^{2+}$-dependent Cy5 blink frequency and average duration of the dark state. The average time spent in the dark state ($\langle\tau_{\text{off}}\rangle$) is close to the time resolution in all cases and is not influenced by $Mg^{2+}$ or $Ca^{2+}$. The frequency at which blinking occurs does not appear to be cation-dependent. All errors correspond to the standard deviation $2\sigma$ as determined from single-exponential fits.

their photostability. However, no significant alterations in response to $M^{2+}$ concentration were observed. The duration of blinking events was characterised by binning the times that fluorophores spent in dark states to histograms ($\tau_{\text{off}}$, see also panels A and F), followed by fitting the cumulative probability distributions 1 - *cumP* to single-exponential decay models. Figures 12.10E and J demonstrate that the rate constants associated with blinking (0.1 s $< \tau <$ 0.9 s) are close to the experimental time resolution of 0.1 s in all cases and they are not significantly altered by the divalent metal ion concentrations assessed herein. Finally, the ratio of the total number of blinking events and the total obvervation time unveils that Cy3 blinks roughly every 1000 s, while Cy5 blinking is slightly less frequent (Figures 12.10E and J).

In conclusion, the excitation intensities used herein ($\sim$ 5 mW) are expected to strike a balance between signal-to-noise ratio and observation time until photobleaching. Furthermore, the average duration and frequency of blinking events is very low due to the combination of OSS and Trolox and appears to be invariable in response to adding $Mg^{2+}$ or $Ca^{2+}$. A previous study that employed the reducing agent β-mercaptoethanol instead of Trolox reported that 80 % of all Cy5 time traces exhibit multiple on/of fluctuations during an observation time of 60 s [393]. As Cy5 blinking leads to anticorrelated changes in intensity in the presence of Cy3, on-off blinking events are indistinguishable from FRET fluctuating between zero and a high value [393]. The rare appearance of acceptor blinking is therefore particularly important in the context of the experimental design of this study. Cy5 also displayed intensity fluctuations that did not reach zero, a previously unprecedented phenomenon. In the context of the FRET experiments performed herein, such blinking events would be expected to lead to intermediate FRET values that are not consistent with the way the experiments were designed [396]. However, alternating excitation of Cy3 and Cy5 (ALEX) is required to unambiguously identify such photophysical artefacts [354, 355, 393].

## 12.3 BOBA FRET source code as of December 2013

BOBA FRET (BOotstrap-Based Analysis of smFRET data) is a script written in Matlab that can be used to quantify the thermodynamic and kinetic cross-sample variability within a set of smFRET data. Bootstrapping is performed in conjunction with four well-established approaches to analyse smFRET time traces in order to estimate the standard deviation associated with the relative occurrence of FRET species (thresholding or multiple Gaussian fit on FRET distributions) or with the transition rates (exponential fit on the dwell-time distributions). The source code is presented in the following sections.

### 12.3.1 bobafret.m (main file)

```
function varargout = bobafret(varargin)
% BOBAFRET - bootstrap-based analysis of smFRET histogram data, version 1.2
% written by Sebastian König, sebastian.koenig@uzh.ch
% Gauss fit algorithm modified from Danny Kowerko
```

```
% danny.kowerko@physik.tu-chemnitz.de
%
% Version 1.1
% Features added by Sebastian König:
% Restructuring the code:  data import, Gaussian fitting and data export are achieved
through M-file functions.
% Data import:  Matrix dimensions are checked.
% listbox_filelist:  number and name of histogram files is displayed.
% Weighting:  histograms can be weighed according to number of counts or not.
% Fitting parameters:  visibility of parameter edit fields changes according to fit type
% Tooltipstrings (rollover explanations) are available.
%
% Version 1.2
% Features added by Sebastian König:
% Menu:  Switching between histogram and dwelltime file import implemented.
% Exponential fitting possible
% Thermodynamic analysis with dwelltime data files
%
% Version 1.2_MH
% Pimped by Mélodie:
% 1) Clustering implemented:  the user has the possibility to cluster the
% transitions.  Clustered data are stored in the "UserData" properties of
% the axes display_results.  ACHTUNG: if the properties "NextPlot" of the
% axes is set to "replace" instead of "replacechildren" and a new plot
% occurs, the "UserData" is reset.
% 2) Corrections of data import:
% calculation of overall FRET histogram when not weighting
% exclusion of the NaN lines in the dwell-time files instead of
% excluding the file itself.
% 3) Re-arrange the code:
% simplify the mmsavedata.m source code
% shorten the name of variables (ex:  handles->h.  eventdata->evd,
% hObject->obj)
% delete unnecessary functions (eval())
% simplify the mmimportdata.m source code regrading the input variable
% (the import algorithm is not anymore selected according to nargin but
% to the pre-defined booleen variables thAna, histDat and threshAna)
% 4) Adapt the title of the loading bar regarding the process occuring
% (before it was always "Gaussian fitting")
% 5) display x-axis, y-axis and title of the axes display_results regarding
% the data plotted
% 6) Bobafret window can be scaled to full screen
% 7) When automatically exporting results, an index is add to the file
% name(s) if files with the same name already exist in the folder
%
% Version 1.3_MH
% Pimped by Mélodie:
% 1) clustering tool take a 4th column in input, containing the file number
% and restoring in output.
% 2) proportions of bobafret tool adapted to any screen dimension
```

```
% 3) 2 transitions by default in clustering tool
% 4) data converion for clusterng input fixed
```

### Initialisation code

```
gui_Singleton = 1;
gui_State = struct('gui_Name', mfilename, ...
     'gui_Singleton', gui_Singleton, ...
     'gui_OpeningFcn', @bobafret_OpeningFcn, ...
     'gui_OutputFcn', @bobafret_OutputFcn, ...
     'gui_LayoutFcn', [] , ...
     'gui_Callback', []);
if nargin && ischar(varargin{1})
     gui_State.gui_Callback = str2func(varargin{1});
end
if nargout
     [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
     gu_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT
```

### Opening function

```
function bobafret_OpeningFcn(obj, evd, h, varargin)
% Executes just before bobafret is made visible.
folder = ";
set(h.pushbutton_currdir, 'String', folder);
set(h.edit_num_hists, 'String', '20');
set(h.edit_num_iter, 'String','20');
set(h.edit_filetype, 'String', 'hist2');

% Create structure with initial fit parameters
fitparam.all = [0 -Inf 0 0 -Inf 0 0 -Inf 0 0 -Inf 0 0 -Inf 0
          0.05 0 0.087 0.05 0.2 0.087 0.05 0.4 0.087 0.05 0.6 0.087 0.05 0.8 0.087
          Inf Inf 0.15 Inf Inf 0.15 Inf Inf 0.15 Inf Inf 0.15 Inf Inf 0.15];
weight = 0;

set(h.edit_folder,'String',folder);

interupt = loading_bar('init', h.bobafret, 45, ...
     'Importing default fitting parameters...');
if ~interupt
     h = guidata(h.bobafret);
     h.barData.prev_var = h.barData.curr_var;
     guidata(h.bobafret, h);

% import start parameters
     varabbr1 = {'amp','cen','wid'};
     varabbr2 = {'lower','start','upper'};
```

```
     for k = 1:3
          for j = 1:3
               for i = 1:5
                    varname = [’h.’  varabbr21,k ’_’ varabbr11,j num2str(i)];
                    varnames1,i = varname;
                    set(eval(varnames1,i), ’String’, fitparam.all(k,(j+3*(i-1))));
                    interupt = loading_bar(’update’, h.bobafret);
               end
          end
     end
     interupt = loading_bar(’close’, h.bobafret);
end

axis(h.display_gof, ’tight’);
[img,map] = imread(’boba.png’);
if ~isempty(map)
     img = ind2rgb(img, map);
end
imagesc(100,100,img, ’Parent’, h.display_gof);
set(h.display_gof,’ytick’,[], ’xtick’, []);

% Choose default command line output for bobafret
h.output = obj;

% Update h structure
guidata(obj, h);

% Normalize the dimensions for scaling
normalizeFig(obj);

function varargout = bobafret_OutputFcn(obj, evd, h)
varargout{1} = h.output;
```

### RMSE analysis

```
function pushbutton_rms_Callback(obj, evd, h)
% Call function mmimportdata
data_type = get(h.radiobutton_histfiles, ’Value’);
[ok, histref, filelist, histfiles, histarray] = mmimportdata(1, data_type, ...
     h.bobafret);
if ok
     set(h.listbox_filelist, ’String’, filelist);
     set(h.edit_numfiles2, ’String’, size(filelist,1));
     set(h.edit_num_hists, ’String’, size(filelist,1));

     % get fitting parameters from edit fields
     fit_type_nr = 5; % import all parameters
     varabbr1 = {’amp’,’cen’,’wid’};
     varabbr2 = {’lower’,’start’,’upper’};
```

```
for k = 1:3
     varname1 = ['fitparam.'  varabbr2{1,k}];
     for j = 1:3
          for i = 1:fit_type_nr
               varnames1{1,k} = varname1;
               varnames1{1,k} = strcat(varnames1{1,k}, '(1,', ...
                    num2str(j+(i-1)*3), ')');
               varname2 = ['h.'  varabbr2{1,k} '_' varabbr1{1,j} num2str(i)];
               value = str2num(get(eval(varname2), 'String'));
               eval([varnames1{1,k} '=value;']);
          end
     end
end

% Gaussian fitting
x_data = histref(:,1);
y_data = histref(:,2);
weight = 0;

[fitres, aic, ex_, cf_, gof] = mmgaussfit(x_data, y_data, fitparam, ...
     fit_type_nr, weight);

% RMSE analysis
set(h.edit_rms_2gaussians, 'String', aic(1,1));
set(h.edit_rms_3gaussians, 'String', aic(1,2));
set(h.edit_rms_4gaussians, 'String', aic(1,3));
set(h.edit_rms_5gaussians, 'String', aic(1,4));

fit_penalty = str2num(get(h.edit_rms_penalty, 'String'));

% Automatically adjust visibility of the fitting parameters
h_3 = [h.lower_wid3 h.start_wid3 h.upper_wid3 h.lower_amp3 h.start_amp3
     h.upper_amp3 h.lower_cen3 h.start_cen3 h.upper_cen3];
h_4 = [h.lower_wid4 h.start_wid4 h.upper_wid4 h.lower_amp4 h.start_amp4
     h.upper_amp4 h.lower_cen4 h.start_cen4 h.upper_cen4];
h_5 = [h.lower_wid5 h.start_wid5 h.upper_wid5 h.lower_amp5 h.start_amp5
     h.upper_amp5 h.lower_cen5 h.start_cen5 h.upper_cen5];

if (fit_penalty * aic(1,2)) >= aic(1,1)
     set(h.num_gauss, 'Value', 2);
     set(h.edit_rms_suggestion, 'String', 2);
     set([h_3 h_4 h_5], 'Visible', 'Off');
elseif (fit_penalty * aic(1,3)) >= aic(1,2)
     set(h.num_gauss, 'Value', 3);
     set(h.edit_rms_suggestion, 'String', 3);
     set(h_3, 'Visible', 'On');
     set([h_4 h_5], 'Visible', 'Off');
elseif (fit_penalty * aic(1,4)) >= aic(1,3)
     set(h.num_gauss, 'Value', 4);
     set(h.edit_rms_suggestion, 'String', 4);
```

```
          set([h_3 h_4],'Visible', 'On');
          set(h_5, 'Visible', 'On');
     else
          set(h.num_gauss, 'Value', 5);
          set(h.edit_rms_suggestion, 'String', 5);
          set([h_3 h_4 h_5], 'Visible', 'On');
     end
     % plot data
     fit_type_nr = get(h.num_gauss, 'Value');
     [fitres, aic, ex_, cf_, gof] = mmgaussfit(x_data, y_data, fitparam, ...
          fit_type_nr, weight);

     h.x_data = histref(:,1);
     h.y_data = histref(:,2);
     bar(h.display_results, h.x_data,h.y_data, 'w');
     set(h.display_results, 'NextPlot', 'add');

     for ii=1:fit_type_nr
          h.y_data = normpdf(histref(:,1), fitres.refb(1,ii), ...
               (fitres.refc(1,ii))/sqrt(2));
          h.y_data(:,1) = h.y_data(:,1)/max(h.y_data(:,1))*fitres.refa(1,ii);
          plot(h.display_results, h.x_data,h.y_data, ...
               'r','LineWidth',3);
     end
     x_label = 'FRET';
     y_label = 'Normalized occurence';
     tl = 'Thermodynamic analysis from FRET distributions';
     set(h.display_results, 'NextPlot', 'replacechildren');
     title(h.display_results, tl);
     xlabel(h.display_results, x_label);
     ylabel(h.display_results, y_label);
     axis(h.display_results, 'normal');
end
```

### Data randomisation and fitting

```
function pushbutton_resfit_Callback(obj, evd, h)
% data import
thAna = strcmp(get(h.submenu_thermodynamics, 'Checked'), 'on');
     if thAna
     % thermodynamic analysis
     histDat = get(h.radiobutton_histfiles, 'Value');
     else
     histDat = 0;
     end
% Question dialog regarding histogram overflow bins
     if histDat == 1
     choice = questdlg('You are about to analyze smFRET histogram files.  Overflow
          bins can drastically reduce the goodness of Gaussian fits.  Would you
```

```
          like to delete the first and the last bin of each histogram?', ...
          'Overflow bins', ...
          'Yes', 'No', 'No');
          switch choice
               case 'Yes'
                    overflowbins = 0;
               case 'No'
                    overflowbins = 1;
          end
     else
     overflowbins = 1;
     end

[ok, histref, filelist, histfiles, histarray, weighting_vector] =
     mmimportdata(thAna, histDat, h.bobafret, overflowbins);

if ok
     set(h.listbox_filelist, 'String', filelist);
     set(h.edit_numfiles2, 'String', size(filelist,1));
     % data randomisation
     % create normalised reference
     num_runs = str2num(get(h.edit_num_iter, 'String'));
     weight = 0;
     fit_type_nr = get(h.num_gauss, 'Value');
     threshAna = get(h.radiobutton_thresholding, 'Value');

     if thAna && histDat % thermodynamic analysis from histogram
          if threshAna % with thresholding
               % get thresholds for threshold-based histogram analysis
               prompt = {'Enter threshold values from lowest to highest.
                    E.g.  "0.22" and "0.65"'};
               if get(h.num_gauss, 'Value') >= 2
               for i=2:get(h.num_gauss, 'Value')
                    prompt{1,i} = ' ';
               end
          end
          thresh_values = [-Inf; Inf; Inf; Inf; Inf; Inf; Inf];
          thresh_values2 = inputdlg(prompt, 'Thresholding');
          for i=1:get(h.num_gauss, 'Value')
               thresh_values((i+1),1) = str2num(thresh_values2{i,1});
          end
          else % with gauss fit
               % get fitting parameters from edit fields
               % Gaussian fitting:  start guesses and boundaries
               varabbr1 = {'amp','cen','wid'};
               varabbr2 = {'lower','start','upper'};
               for k = 1:3
                    varname1 = ['fitparam.'  varabbr2{1,k}];
                    for j = 1:3
                         for i = 1:fit_type_nr
```

```
               varnames1{1,k} = varname1;
               varnames1{1,k} = strcat(varnames1{1,k}, ’(1,’, ...
                  num2str(j+(i-1)*3), ’)’);
               varname2 = [’h.’  varabbr2{1,k} ’_’ varabbr1{1,j}
                  num2str(i)];
               value = str2num(get(eval(varname2), ’String’));
               eval([varnames1{1,k} ’=value;’]);
            end
         end
      end
   end
elseif ∼thAna % kinetic analysis
   % exponential fitting:  start guesses and boundaries
   if get(h.y_offset, ’Value’) == 0;
      % equivalent to no y-offset
      fitparam.lower(1,1) = 0;
      fitparam.start(1,1) = 0;
      fitparam.upper(1,1) = 0;
   else
      fitparam.lower(1,1) = str2num(get(h.edit_lower_y_offset, ’String’));
      fitparam.start(1,1) = str2num(get(h.edit_start_y_offset, ’String’));
      fitparam.upper(1,1) = str2num(get(h.edit_upper_y_offset, ’String’));
   end
   if get(h.x_offset, ’Value’) == 0;
      % equivalent to no x-offset
      fitparam.lower(1,2) = 0;
      fitparam.start(1,2) = 0;
      fitparam.upper(1,2) = 0;
   else
      fitparam.lower(1,2) = str2num(get(h.edit_lower_x_offset, ’String’));
      fitparam.start(1,2) = str2num(get(h.edit_start_x_offset, ’String’));
      fitparam.upper(1,2) = str2num(get(h.edit_upper_x_offset, ’String’));
   end
   % get values from edit fields
   if fit_type_nr <= 4 % single-/multiexp.  fit
      varabbr1 = {’amp’,’cen’}; % here:  cen = time constant of the decay
      varabbr2 = {’lower’,’start’,’upper’};
      for k = 1:3
         varname1 = [’fitparam.’  varabbr2{1,k}];
         for j = 1:2
            for i = 1:fit_type_nr
               varnames1{1,k} = varname1;
               varnames1{1,k} = strcat(varnames1{1,k}, ’(1,’, ...
                  num2str(j+(i-1)*2+2), ’)’);
               varname2 = [’h.’  varabbr2{1,k} ’_’ varabbr1{1,j}
                  num2str(i)];
               value = str2num(get(eval(varname2), ’String’));
               eval([varnames1{1,k} ’=value;’]);
            end
         end
```

```
          end
     else % stretched exponential fit
          varabbr1 = {’amp’,’cen’}; % here:  cen = time constant of the decay
          varabbr2 = {’lower’,’start’,’upper’};
          for k = 1:3
               varname1 = [’fitparam.’  varabbr2{1,k}];
               for j = 1:2
                    varnames1{1,k} = varname1;
                    varnames1{1,k} = strcat(varnames1{1,k}, ’(1,’, ...
                         num2str(j+2), ’)’);
                    varname2 = [’h.’  varabbr2{1,k} ’_’ varabbr1{1,j} num2str(1)];
                    value = str2num(get(eval(varname2), ’String’));
                    eval([varnames1{1,k} ’=value;’]);
               end
          end
          % import values for beta
          fitparam.lower(1,5) = str2num(get(h.lower_amp2, ’String’));
          fitparam.start(1,5) = str2num(get(h.start_amp2, ’String’));
          fitparam.upper(1,5) = str2num(get(h.upper_amp2, ’String’));
     end
end

% randomly select histograms, exponential/Gaussian fitting

num_hist = str2num(get(h.edit_num_hists, ’String’));
if thAna && histDat % thermodynamic analysis from histogram files
     % initialise loading bar
     if threshAna
          str = ’thresholding’;
     else
          str = ’Gaussian fitting’;
     end
     interupt = loading_bar(’init’, h.bobafret, num_runs, ...
          [’Performing randomisation and ’ str ’...’]);
     if ~interupt
          h = guidata(h.bobafret);
          h.barData.prev_var = h.barData.curr_var;
          guidata(h.bobafret, h);
          weighting = get(h.radiobutton_weighing_yes, ’Value’);
          if weighting == 1
               for j = 1:size(histarray,2)
                    histarray{1,j}(:,2) = histarray{1,j}(:,2)/
                         sum(histarray{1,j}(:,2));
               end
          else
               weighting_vector = ones(size(histfiles,1),1);
          end
          for k = 1:str2double(get(h.edit_num_iter, ’String’))
               % update loading bar
               interupt = loading_bar(’update’, h.bobafret);
```

```
                randomhists = randsample(size(histfiles,1),num_hist,true, ...
                    weighting_vector);
                histall = histarray{1,randomhists(1)};
                for i = 2:num_hist
                    histcurr = histarray{1,randomhists(i)};
                    histall(:,2:size(histall,2)) = (histall(:, ...
                        2:size(histall,2)) + histcurr(:,2:size(histall,2)));
                end
                histall(:,2) = histall(:,2)/sum(histall(:,2));
                if ~threshAna % with gauss fit
                    % start Gaussian fitting
                    x_data = histall(:,1);
                    y_data = histall(:,2);
                    [fitres, aic, ex_, cf_, gof] = mmgaussfit(x_data, ...
                        y_data, fitparam, fit_type_nr, weight);
                    % get fitting results
                    coeffs = coeffvalues(cf_);
                    fitres2.adj_s(k,1) = gof.adjrsquare;
                    for gg=1:fit_type_nr
                        fitres2.a(k,gg) = coeffs(1,3*gg-2); % Gauss amplitude
                        fitres2.b(k,gg) = coeffs(1,3*gg-1); % Gauss center
                        fitres2.c(k,gg) = coeffs(1,3*gg-0); % Gauss width
                    end
                else % with thresholding
                    for i = 1:(get(h.num_gauss, 'Value')+1)
                        results.population(k,i) = sum(histall((histall(:,1) >
                            thresh_values(i,1) & histall(:,1) <=
                            thresh_values(i+1)),2));
                    end
                end
            end
            % close loading bar
            interupt = loading_bar('close', h.bobafret);
        end
    else
        % adjust number of replicates for bootstrapping y/n
        number_replicates = size(histarray,2);
        if number_replicates ~= str2num(get(h.edit_num_hists, 'String'))
            question = ['Number of molecules with relevant dwell times:  ', ...
                num2str(number_replicates), ...
                '.  Should the number of replicates be adjusted in the
                    resampling process (suggested)?'];
            choice = questdlg(question, 'Adjust number of replicates', ...
                'Yes','No','Cancel','Cancel');
            switch choice
                case 'Yes'
                    set(h.edit_num_hists, 'String', ...
                        num2str(number_replicates));
                case 'Cancel'
                    return;
```

```
     end
end
if thAna % thermodynamic analysis from dwell-times
     num_hist = str2num(get(h.edit_num_hists, ’String’));
     interupt = loading_bar(’init’, h.bobafret, num_runs, ...
          ’Performing randomisation and population calculation...’);
     if ∼interupt
          h = guidata(h.bobafret);
          h.barData.prev_var = h.barData.curr_var;
          guidata(h.bobafret, h);
          for k = 1:str2double(get(h.edit_num_iter, ’String’))
               % update loading bar
               interupt = loading_bar(’update’, h.bobafret);
               randomhists = randsample(size(histarray,2),num_hist, ...
                    true, weighting_vector);
               transitions = [];
               for j = 1:num_hist
                    if isempty(find(isnan(histarray{randomhists(j,1)})))
                              == 1
                         transitions = [transitions; histarray{1, ...
                         randomhists(j,1)}];
                    else
                         indexing(1,(size(indexing,2)+1)) = j;
                    end
               end
               transitions = unique(transitions(:, ...
                    (size(transitions,2)-1):size(transitions,2)), ’rows’);
               histall = [];
               histall(:,2:3) = transitions;
               for i = 1:num_hist
                    histarray{1,randomhists(i,1)} = sortrows(
                         histarray{1,randomhists(i,1)},2);
                    [C,ia,ic] = unique(histarray{1,randomhists(i,1)}(:,2));
                    if size(histarray{1,randomhists(i,1)}) == size(histall)
                         histall(:,1) = histall(:,1) + histarray{1, ...
                              randomhists(i,1)}(:,1);
                    else
                         for l = 1:size(histarray{1,randomhists(i,1)},1)
                         j = 1;
                         while isequal(histarray{1,randomhists(i,1)}
                              (l,2), histall(j,2)) == 0
                              j = j + 1;
                         end
                         histall(j,1) = histall(j,1) + histarray{1, ...
                              randomhists(i,1)}(l,1);
                    end
               end
          end
          for i = 1:size(transitions,1)
               results.population(k,i) = histall(i,1)/sum(histall(:,1));
```

```
          end
     end
     % close loading bar
     interupt = loading_bar(’close’, h.bobafret);
end
else % kinetic analysis
     % initialise loading bar
     interupt = loading_bar(’init’, h.bobafret, num_runs, ...
          ’Performing randomisation and exponential fitting...’);
     if ~interupt
          h = guidata(h.bobafret);
          h.barData.prev_var = h.barData.curr_var;
          guidata(h.bobafret, h);
          for k = 1:str2double(get(h.edit_num_iter, ’String’))
               % update loading bar
               interupt = loading_bar(’update’, h.bobafret);
               randomhists = randsample(size(histarray,2),num_hist, ...
                    true, weighting_vector);
               histall = [];
               for i = 1:size(randomhists,1)
                    histall = [histall; histarray{1,randomhists(i,1)}];
               end
               histall = histall(:,1);
               % sort dwell times
               histall = sort(histall);
               histall(:,2) = 1 - linspace(1, size(histall,1), ...
                    size(histall,1))/size(histall,1);
               [C,ia,ic] = unique(histall(:,1));
               histall = histall(ia,:);
               histall = [0, 1; histall];
               % lauch exponential fitting
               x_data = histall(:,1);
               y_data = histall(:,2);
               [fitres,ex_, cf_, gof] = mmexpfit(x_data, y_data, ...
                    fitparam,fit_type_nr);
               % get fitting results
               fitres2.adj_s(k,1) = gof.adjrsquare;
               if fit_type_nr == 5
                    fitres2.a(k,:)  = fitres.refa;
                    % stretched exponential fit amplitude
                    fitres2.c(k,:)  = fitres.refc;
                    % stretched exponential fit decay constant
                    fitres2.xo(k,:)  = fitres.refxo;
                    % stretched exponential fit x-offset
                    fitres2.yo(k,:)  = fitres.refyo;
                    % stretched exponential fit y-offset
                    fitres2.beta(k,:)  = fitres.refbeta;
                    % stretched exponential fit beta
               else
                    fitres2.a(k,:)  = fitres.refa;
```

```
                         % Exponential fits, amplitudes
                         fitres2.c(k,:)  = fitres.refc;
                         % Exponential fits, decay constants
                         fitres2.xo(k,:)  = fitres.refxo;
                         % Exponential fits, x-offset
                         fitres2.yo(k,:)  = fitres.refyo;
                         % Exponential fits, y-offset
                    end
               end
               % close loading bar
               interupt = loading_bar(’close’, h.bobafret);
          end
     end
end
% fit reference histogram
if thAna && histDat && ~threshAna
     % thermodynamic analysis from histograms with thresholding
     x_data = histref(:,1);
     y_data = histref(:,2);
     [fitres, aic, ex_, cf_, gof] = mmgaussfit(x_data, y_data, fitparam, ...
     fit_type_nr,weight);
     fitres2.refa = fitres.refa; % Gaussparameter(s) amplitude
     fitres2.refb = fitres.refb; % Gaussparameter(s) center
     fitres2.refc = fitres.refc; % Gaussparameter(s) width
elseif ~thAna % kinetic analysis
     x_data = histref(:,1);
     y_data = histref(:,2);
     [fitres, ex_, cf_, gof] = mmexpfit(x_data,y_data,fitparam,fit_type_nr);
     if fit_type_nr == 5
          fitres2.refa = fitres.refa ;
          % stretched exponential fit amplitude
          fitres2.refc = fitres.refc;
          % stretched exponential fit decay constant
          fitres2.refxo = fitres.refxo;
          % stretched exponential fit x-offset
          fitres2.refyo = fitres.refyo;
          % stretched exponential fit y-offset
          fitres2.refbeta = fitres.refbeta;
          % stretched exponential fit beta
     else
          fitres2.refa = fitres.refa; % Exponential fits, amplitudes
          fitres2.refc = fitres.refc; % Exponential fits, decay constants
          fitres2.refxo = fitres.refxo; % Exponential fits, x-offset
          fitres2.refyo = fitres.refyo; % Exponential fits, y-offset
     end
end
% sort data
if thAna % thermodynamic analysis
     if histDat % from histograms
          if ~threshAna % with gauss fit
```

```
                fitres2.amplitudessorted = sort(fitres2.a);
                fitres2.centersorted = sort(fitres2.b);
                fitres2.widthsorted = sort(fitres2.c);
            else % with thresholding
                for i = 1:(get(h.num_gauss, 'Value')+1)
                    results.states{1,i} = [num2str(thresh_values(i,1)), ...
                        ' to ', num2str(thresh_values((i+1),1)), '; '];
                end
                for j = 1:(get(h.num_gauss, 'Value')+1)
                    results.reference(1,j) = sum(histref(thresh_values(j,1) <
                        histref(:,1) & histref(:,1) <= thresh_values(j+1),2));
                end
            end
        else % from dwell-times
            for i = 1:  size(histall,1)
                results.states(1,i) = {[num2str(histall(i, ...
                    (size(histall,2)-1):size(histall,2))), ' FRET']};
            end
            results.reference(1,:)  = histref(:,1)/sum(histref(:,1));
            sort(results.population);
        end
    else % kinetic analysis
        if fit_type_nr == 5
            fitres2.sorted = sortrows([fitres2.beta, fitres2.c, ...
            fitres2.yo, fitres2.xo, fitres2.a], 1);
        else
            fitres2.sorted = sortrows([fitres2.c, fitres2.yo, ...
            fitres2.xo, fitres2.a], 1);
        end
    end
    % numerical integration, ref
    if thAna && histDat % thermodynamic analysis from histograms
        if ~threshAna
            for ii=1:fit_type_nr
                fitres2.distref{ii}(:,1) = normpdf(histall(:,1), ...
                fitres2.refb(1,ii), fitres2.refc(1,ii));
                fitres2.distref{ii}(:,1) = fitres2.distref{ii}(:,1)/ ...
                    max(fitres2.distref{ii}(:,1))*fitres2.refa(1,ii);
            end
            % relative occurrence, ref
            for ii = 1:fit_type_nr
                fitres2.reloccurrenceref(1,ii) = sum(fitres2.distref{ii}(:, ...
                    1))/sum(cellfun(@sum, fitres2.distref));
            end
            % numerical integration, bootstrap samples
            for k= 1:num_runs
                for ii = 1:fit_type_nr
                    fitres2.dist{ii}(:,k) = normpdf(histall(:,1), ...
                        fitres2.b(k,ii), fitres2.c(k,ii));
                    fitres2.dist{ii}(:,k) = fitres2.dist{ii}(:,k)/ ...
```

```
                    max(fitres2.dist{ii}(:,k))∗fitres2.a(k,ii);
                end
            end
            % relative occurrence, boostrap samples
            fitres2.huminttemp = cellfun(@sum, fitres2.dist, ...
                ’UniformOutput’,false);
            fitres2.numint = fitres2.numinttemp{1};
            for ii = 2:fit_type_nr
                fitres2.numint = fitres2.numint + fitres2.numinttemp{ii};
            end
            for ii = 1:fit_type_nr
                fitres2.reloccurrence(:,ii) = sum(fitres2.dist{ii}(:,:))./ ...
                    fitres2.numint;
            end
            fitres2.reloccurrencesorted=sort(fitres2.reloccurrence);
        end
    end
    % save and export data
    if thAna && histDat && ∼threshAna
        % thermodynamic analysis from histogram by gauss fit
        results.fitres2 = fitres2;
    elseif ∼thAna % kinetic analysis
        results.fitres2 = fitres2;
        results.fit_type_nr = fit_type_nr;
        results.x_offset = get(h.x_offset, ’Value’);
        results.y_offset = get(h.y_offset, ’Value’);
    end
    % plot data
    if thAna % thermodynamic analysis
        if histDat % from histograms
            h.x_data = histref(:,1);
            h.y_data = histref(:,2);
            bar(h.display_results, h.x_data,h.y_data,’w’);
            set(h.display_results, ’NextPlot’, ’add’);
            if ∼threshAna % with gauss fit
                for ii=1:fit_type_nr
                    h.y_data(:,ii) = normpdf(histall(:,1), ...
                        fitres2.refb(1,ii), (fitres2.refc(1,ii))/sqrt(2));
                    h.y_data(:,ii) = h.y_data(:,ii)/ ...
                    max(h.y_data(:,ii))∗fitres2.refa(1,ii);
                end
                switch fit_type_nr
                    case {2}
                        plot(h.display_results, h.x_data, ...
                            h.y_data(:,1), ’Color’, [0 0 0.9], ...
                            ”LineWidth’, 3);
                        plot(h.display_results, h.x_data, ...
                            h.y_data(:,2), ’Color’, [0.5 0.5 1], ...
                            ”LineWidth’, 3);
                    case {3}
```

```
          plot(h.display_results, h.x_data, ...
               h.y_data(:,1), ’Color’, [0 0 0.8], ...
               ”LineWidth’, 3);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,2), ’Color’, [0.2 0.2 1], ...
               ”LineWidth’, 3);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,3), ’Color’, [0.6 0.6 1], ...
               ”LineWidth’, 3);
     case {4}
          plot(h.display_results, h.x_data, ...
               h.y_data(:,1), ’Color’, [0 0 0.6], ...
               ”LineWidth’, 3);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,2), ’Color’, [0 0 1], ...
               ”LineWidth’, 3);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,3), ’Color’, [0.4 0.4 1], ...
               ”LineWidth’, 3);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,4), ’Color’, [0.8 0.8 1], ...
               ”LineWidth’, 3);
     otherwise
          plot(h.display_results, h.x_data, ...
               h.y_data(:,1), ’Color’, [0 0 0.6], ...
               ”LineWidth’, 3);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,2), ’Color’, [0 0 0.9], ...
               ”LineWidth’, 3);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,3), ’Color’, [0.2 0.2 1], ...
               ”LineWidth’, 3);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,4), ’Color’, [0.5 0.5 1], ...
               ”LineWidth’, 3);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,5), ’Color’, [0.8 0.8 1], ...
               ”LineWidth’, 3);
end
for ii=1:fit_type_nr
     h.y_data(:,ii) = normpdf(histall(:,1), ...
          fitres2.refb(1,ii),(fitres2.widthsorted(1, ...
          ii))/sqrt(2));
     h.y_data(:,ii) = h.y_data(:,ii)/ ...
          max(h.y_data(:,ii))∗fitres2.amplitudessorted(1,ii);
end
switch fit_type_nr
case {2}
          plot(h.display_results, h.x_data, ...
               h.y_data(:,1),’-’,’Color’,[0 0 0.9],”LineWidth’,2);
```

```
          plot(h.display_results, h.x_data, ...
               h.y_data(:,2), ’-’, ’Color’, [0.5 0.5 1],...
               ’LineWidth’, 2);
     case {3}
          plot(h.display_results, h.x_data, ...
               h.y_data(:,1), ’-’, ’Color’, [0 0 0.8], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,2), ’-’, ’Color’, [0.2 0.2 1], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,3), ’-’, ’Color’, [0.6 0.6 1], ...
               ’LineWidth’, 2);
     case {4}
          plot(h.display_results, h.x_data, ...
               h.y_data(:,1), ’-’, ’Color’, [0 0 0.6], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,2), ’-’, ’Color’, [0 0 1], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,3), ’-’, ’Color’, [0.4 0.4 1], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,4), ’-’, ’Color’, [0.8 0.8 1], ...
               ’LineWidth’, 2);
     otherwise
          plot(h.display_results, h.x_data, ...
               h.y_data(:,1), ’-’, ’Color’, [0 0 0.6], ...
               ’LineWidth’, 2);
               plot(h.display_results, h.x_data, ...
               h.y_data(:,2), ’-’, ’Color’, [0 0 0.9], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,3), ’-’, ’Color’, [0.2 0.2 1], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,4), ’-’, ’Color’, [0.5 0.5 1], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,5), ’-’, ’Color’, [0.8 0.8 1], ...
               ”LineWidth’, 2);
end
for ii=1:fit_type_nr
     h.y_data(:,ii) = normpdf(histall(:,1), ...
          fitres2.refb(1,ii),(fitres2.widthsorted(
          size(fitres2.centersorted,1),ii))/sqrt(2));
     h.y_data(:,ii) = h.y_data(:,ii)/max(h.y_data(:,ii))* ...
          fitres2.amplitudessorted(
          size(fitres2.centersorted,1),ii);
```

```
end
switch fit_type_nr
     case {2}
          plot(h.display_results, h.x_data, ...
               h.y_data(:,1),’-’, ’Color’, [0 0 0.9], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,2), ’-’, ’Color’, [0.5 0.5 1], ...
               ’LineWidth’, 2);
     case {3}
          plot(h.display_results, h.x_data, ...
               h.y_data(:,1), ’-’, ’Color’, [0 0 0.8], ...
               ”LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
                    h.y_data(:,2),’-’, ’Color’, [0.2 0.2 1], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,3), ’-’, ’Color’, [0.6 0.6 1], ...
               ’LineWidth’, 2);
     case {4}
          plot(h.display_results, h.x_data, ...
               h.y_data(:,1), ’-’, ’Color’, [0 0 0.6], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
                    h.y_data(:,2), ’-’, ’Color’, [0 0 1], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,3), ’-’, ’Color’, [0.4 0.4 1], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,4), ’-’, ’Color’, [0.8 0.8 1], ...
               ’LineWidth’, 2);
     otherwise
          plot(h.display_results, h.x_data, ...
                    h.y_data(:,1),’-’, ’Color’, [0 0 0.6], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,2), ’-’, ’Color’, [0 0 0.9], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,3), ’-’, ’Color’, [0.2 0.2 1], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,4), ’-’, ’Color’, [0.5 0.5 1], ...
               ’LineWidth’, 2);
          plot(h.display_results, h.x_data, ...
               h.y_data(:,5), ’-’, ’Color’, [0.8 0.8 1], ...
               ’LineWidth’, 2);
end
set(h.display_results, ’NextPlot’, ’replacechildren’);
```

```
          x_label = ’FRET’;
          y_label = ’Normalized occurence’;
          tl = ’Thermodynamic analysis from FRET histograms’;
          title(h.display_results, tl);
          xlabel(h.display_results, x_label);
          ylabel(h.display_results, y_label);
          axis(h.display_results, ’normal’);
          axis(h.display_gof, ’tight’);
          h.y_data = fitres2.adj_s(:,1);
          hist(h.display_gof, h.y_data,20);
          title(h.display_gof, ’Goodness of fit’);
          xlabel(h.display_gof, ’adjusted Rˆ2’);
          ylabel(h.display_gof, ’counts’);
          else % with thresholding
               for i=1:get(h.num_gauss, ’Value’)
                    plot(h.display_results, [thresh_values((i+1),1)
                         thresh_values((i+1),1)], ...
                         [0 max(histref(:,2))], ’r’, ”LineWidth’, 2);
               end
               set(h.display_results, ’NextPlot’, ’replacechildren’);
               % display message box
               msgbox(’Done.  Results saved in file
               "results_randomisation.txt".’, ...
                    ’Analysis completed’, ’help’);
          end
     else % from dwell-times
          set(h.display_results, ’NextPlot’, ’replacechildren’);
          % display message box
          x_label = ’FRET’;
          y_label = ’Normalized occurence’;
          tl = ’Thermodynamic analysis from dwell-times’;
          title(h.display_results, tl);
          xlabel(h.display_results, x_label);
          ylabel(h.display_results, y_label);
     end
else % kinetic analysis
% plot raw data
h.x_data = histref(:,1);
h.y_data = histref(:,2);
plot(h.display_results, h.x_data,h.y_data, ’+’, ’LineWidth’, 5);
set(h.display_results, ’NextPlot’, ’add’);
% plot fits for reference and bootstrapped data
if fit_type_nr == 5 % stretched exponential fit
     plot_ref = fitres2.refyo + fitres2.refa* ...
          exp(-((x_data-fitres2.refxo)/ ...
          fitres2.refc).ˆfitres2.refbeta);
     plot_bootstrap1 = fitres2.sorted(1,3) + fitres2.sorted(1,5)* ...
          exp(-((x_data-fitres2.sorted(1,4))/fitres2.sorted(1,2)).ˆ ...
          fitres2.sorted(1,1));
     plot_bootstrap2 = fitres2.sorted(size(fitres2.sorted,1),3) + ...
```

```
               fitres2.sorted(size(fitres2.sorted,1),5)* ...
               exp(-((x_data-fitres2.sorted(size(fitres2.sorted,1),4))/ ...
               fitres2.sorted(size(fitres2.sorted,1),2)).^ ...
               fitres2.sorted(size(fitres2.sorted,1),1));
          plot(h.display_results, h.x_data,plot_ref, ’r’, ’LineWidth’, 2);
          plot(h.display_results, h.x_data,plot_bootstrap1,’-’, ’Color’, ...
               [1 0 0],’LineWidth’, 1.5);
          plot(h.display_results, h.x_data,plot_bootstrap2,’-’, ’Color’, ...
               [1 0 0],’LineWidth’, 1.5);
          choice = questdlg(’This graph illustrates both the variation in
               the decay constant and the stretching factor beta.’, ...
               ’Information’, ...
          ’Good to know.’, ’Good to know.’);
     else % single-/multiexponential fit
          plot_ref = fitres2.refyo;
          plot_bootstrap1 = fitres2.sorted(1,(1 + fit_type_nr));
          plot_bootstrap2 = fitres2.sorted(size(fitres2.sorted,1), ...
               (1 + fit_type_nr));
          for i = 1:fit_type_nr
               plot_ref = plot_ref + fitres2.refa(1,i)* ...
                    exp(-(x_data-fitres2.refxo)/fitres2.refc(1,i));
               plot_bootstrap1 = plot_bootstrap1 + ...
                    fitres2.sorted(1,(fit_type_nr+2+i))* ...
                    exp(-(x_data-fitres2.sorted(1,(2+fit_type_nr)))/ ...
                    fitres2.sorted(1,i));
               plot_bootstrap2 = plot_bootstrap2 + ...
                    fitres2.sorted(size(fitres2.sorted,1),(fit_type_nr+2+ ...
                    i))*exp(-(x_data-fitres2.sorted(size(fitres2.sorted,1), ...
                    (2+fit_type_nr)))/fitres2.sorted(size(fitres2.sorted,1),i));
          end
          plot(h.display_results, h.x_data,plot_ref, ’r’, ”LineWidth’, 2);
          plot(h.display_results, h.x_data,plot_bootstrap1,’-’, ...
               ’Color’,[1 0 0],”LineWidth’, 1.5);
          plot(h.display_results, h.x_data,plot_bootstrap2,’-’, ...
               ’Color’,[1 0 0],”LineWidth’, 1.5);
     end
     set(h.display_results, ’NextPlot’, ’replacechildren’);
     x_label = ’dwell-times (s)’;
     y_label = ’normalized (1 - cum(P))’;
     tl = ’Kinetic analysis from dwell-times’;
     title(h.display_results, tl);
     xlabel(h.display_results, x_label);
     ylabel(h.display_results, y_label);
end
if get(h.radiobutton_weighing_yes, ’Value’) == 1
     if ~thAna
          choice = questdlg(’You have performed weighted bootstrapping,
               i.e.  time traces with longer observation times are more
               likely to be selected.  This will naturally favor the selection
               of longer dwell times and may may lead to deviations
```

```
                    from the reference (blue dots, solid line).', ...
               'Information', ...
               'Good to know.', 'Good to know.');
          end
     end
     choice = questdlg('Would you like to export the results to the current
          Matlab folder?', ...
          'export results', ...
          'Yes', 'No', 'Yes');
     switch choice
          case 'Yes'
               mmsavedata([thAna, histDat, threshAna], results, h.display_results);
          case 'No'
     end
end
```

### graphical user interface (GUI)

```
function edit_folder_Callback(obj, evd, h)
% Edit buttons for data import and bootstrapping
function edit_folder_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

function pushbutton_currdir_Callback(obj, evd, h)
currdir_str = get(h.edit_folder, 'String');
if get(h.radiobutton_histfiles, 'Value') && ~strcmp(get(h.submenu_kinetics, ...
          'Checked'), 'On')
     str = 'Choose directory containing histogram files:';
else
     str = 'Choose directory containing dwell time files:';
end
folder = uigetdir(currdir_str, str);
if folder
     set(h.edit_folder,'String',folder);
end

function edit_num_hists_Callback(obj, evd, h)
function edit_num_hists_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
     'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

function edit_num_iter_Callback(obj, evd, h)
function edit_num_iter_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
```

```
    'defaultUicontrolBackgroundColor'))
    set(obj,'BackgroundColor','white');
end

function num_gauss_Callback(obj, evd, h)
% ...adjust visibilities...
num_gauss = get(h.num_gauss, 'Value');
if num_gauss == 1
    tick2 = strcmp(get(h.submenu_kinetics, 'Checked'),'On');
    if tick2 == 1
        set(h.lower_wid2, 'Visible', 'Off');
        set(h.start_wid2, 'Visible', 'Off');
        set(h.upper_wid2, 'Visible', 'Off');
        set(h.lower_amp2, 'Visible', 'Off');
        set(h.start_amp2, 'Visible', 'Off');
        set(h.upper_amp2, 'Visible', 'Off');
        set(h.lower_cen2, 'Visible', 'Off');
        set(h.start_cen2, 'Visible', 'Off');
        set(h.upper_cen2, 'Visible', 'Off');

        set(h.lower_wid3, 'Visible', 'Off');
        set(h.start_wid3, 'Visible', 'Off');
        set(h.upper_wid3, 'Visible', 'Off');
        set(h.lower_amp3, 'Visible', 'Off');
        set(h.start_amp3, 'Visible', 'Off');
        set(h.upper_amp3, 'Visible', 'Off');
        set(h.lower_cen3, 'Visible', 'Off');
        set(h.start_cen3, 'Visible', 'Off');
        set(h.upper_cen3, 'Visible', 'Off');

        set(h.lower_cen4, 'Visible', 'Off');
        set(h.start_cen4, 'Visible', 'Off');
        set(h.upper_cen4, 'Visible', 'Off');
        set(h.lower_wid4, 'Visible', 'Off');
        set(h.start_wid4, 'Visible', 'Off');
        set(h.upper_wid4, 'Visible', 'Off');
        set(h.lower_amp4, 'Visible', 'Off');
        set(h.start_amp4, 'Visible', 'Off');
        set(h.upper_amp4, 'Visible', 'Off');

        set(h.lower_wid5, 'Visible', 'Off');
        set(h.start_wid5, 'Visible', 'Off');
        set(h.upper_wid5, 'Visible', 'Off');
        set(h.lower_amp5, 'Visible', 'Off');
        set(h.start_amp5, 'Visible', 'Off');
        set(h.upper_amp5, 'Visible', 'Off');
        set(h.lower_cen5, 'Visible', 'Off');
        set(h.start_cen5, 'Visible', 'Off');
        set(h.upper_cen5, 'Visible', 'Off');
```

```
        set(h.amp2, 'String', 'amplitude 2', 'TooltipString', ...
            '<html><b>Amplitude</b><br/>Initial occurrence at t = 0.  Adopts
                a value of 1<br/>for normalised cumulative probability
                distributions</html>');
        set(h.display_eq, 'String', 'y = a*exp{-((x-b)/tau)} + c');
    else
        set(h.amp2, 'String', 'amplitude 2', 'TooltipString', ...
            '<html><b>Amplitude</b><br/>Defines the amplitude ("height")
                of the Gaussian fit function.<br/>See also "A" in equation (2)
                of the original publication.</html>');
        set(h.lower_wid2, 'Visible', 'On');
        set(h.start_wid2, 'Visible', 'On');
        set(h.upper_wid2, 'Visible', 'On');
        set(h.lower_amp2, 'Visible', 'On', 'String', '0');
        set(h.start_amp2, 'Visible', 'On', 'String', '0.05');
        set(h.upper_amp2, 'Visible', 'On', 'String', 'Inf');
        set(h.lower_cen2, 'Visible', 'On');
        set(h.start_cen2, 'Visible', 'On');
        set(h.upper_cen2, 'Visible', 'On');

        set(h.lower_wid3, 'Visible', 'Off');
        set(h.start_wid3, 'Visible', 'Off');
        set(h.upper_wid3, 'Visible', 'Off');
        set(h.lower_amp3, 'Visible', 'Off');
        set(h.start_amp3, 'Visible', 'Off');
        set(h.upper_amp3, 'Visible', 'Off');
        set(h.lower_cen3, 'Visible', 'Off');
        set(h.start_cen3, 'Visible', 'Off');
        set(h.upper_cen3, 'Visible', 'Off');

        set(h.lower_cen4, 'Visible', 'Off');
        set(h.start_cen4, 'Visible', 'Off');
        set(h.upper_cen4, 'Visible', 'Off');
        set(h.lower_wid4, 'Visible', 'Off');
        set(h.start_wid4, 'Visible', 'Off');
        set(h.upper_wid4, 'Visible', 'Off');
        set(h.lower_amp4, 'Visible', 'Off');
        set(h.start_amp4, 'Visible', 'Off');
        set(h.upper_amp4, 'Visible', 'Off');

        set(h.lower_wid5, 'Visible', 'Off');
        set(h.start_wid5, 'Visible', 'Off');
        set(h.upper_wid5, 'Visible', 'Off');
        set(h.lower_amp5, 'Visible', 'Off');
        set(h.start_amp5, 'Visible', 'Off');
        set(h.upper_amp5, 'Visible', 'Off');
        set(h.lower_cen5, 'Visible', 'Off');
        set(h.start_cen5, 'Visible', 'Off');
        set(h.upper_cen5, 'Visible', 'Off');
    end
```

```
elseif num_gauss == 2

     tick2 = strcmp(get(h.submenu_kinetics, 'Checked'),'On');
     if tick2 == 1
          set(h.amp2, 'String', 'amplitude 2', 'TooltipString', ...
               '<html><b>Amplitude</b><br/>Initial occurrence at t = 0.  Adopts
                    a value of 1<br/>for normalised cumulative probability
                    distributions</html>');
          set(h.display_eq, 'String', ...
               'y = a1*exp{-((x-b1)/tau1)} + a2*exp{-((x-b2)/tau2)} + c');
          set(h.lower_wid2, 'Visible', 'Off');
          set(h.start_wid2, 'Visible', 'Off');
          set(h.upper_wid2, 'Visible', 'Off');
          set(h.lower_amp2, 'Visible', 'On', 'String', '0.01');
          set(h.start_amp2, 'Visible', 'On', 'String', '0.8');
          set(h.upper_amp2, 'Visible', 'On', 'String', '1');
          set(h.lower_cen2, 'Visible', 'On');
          set(h.start_cen2, 'Visible', 'On');
          set(h.upper_cen2, 'Visible', 'On');

          set(h.lower_wid3, 'Visible', 'Off');
          set(h.start_wid3, 'Visible', 'Off');
          set(h.upper_wid3, 'Visible', 'Off');
          set(h.lower_amp3, 'Visible', 'Off');
          set(h.start_amp3, 'Visible', 'Off');
          set(h.upper_amp3, 'Visible', 'Off');
          set(h.lower_cen3, 'Visible', 'Off');
          set(h.start_cen3, 'Visible', 'Off');
          set(h.upper_cen3, 'Visible', 'Off');
          set(h.lower_cen4, 'Visible', 'Off');
          set(h.start_cen4, 'Visible', 'Off');
          set(h.upper_cen4, 'Visible', 'Off');
          set(h.lower_wid4, 'Visible', 'Off');
          set(h.start_wid4, 'Visible', 'Off');
          set(h.upper_wid4, 'Visible', 'Off');
          set(h.lower_amp4, 'Visible', 'Off');
          set(h.start_amp4, 'Visible', 'Off');
          set(h.upper_amp4, 'Visible', 'Off');

          set(h.lower_wid5, 'Visible', 'Off');
          set(h.start_wid5, 'Visible', 'Off');
          set(h.upper_wid5, 'Visible', 'Off');
          set(h.lower_amp5, 'Visible', 'Off');
          set(h.start_amp5, 'Visible', 'Off');
          set(h.upper_amp5, 'Visible', 'Off');
          set(h.lower_cen5, 'Visible', 'Off');
          set(h.start_cen5, 'Visible', 'Off');
          set(h.upper_cen5, 'Visible', 'Off');
     else
```

```
        set(h.lower_wid2, ’Visible’, ’On’);
        set(h.start_wid2, ’Visible’, ’On’);
        set(h.upper_wid2, ’Visible’, ’On’);
        set(h.lower_amp2, ’Visible’, ’On’, ’String’, ’0’);
        set(h.start_amp2, ’Visible’, ’On’, ’String’, ’0.05’);
        set(h.upper_amp2, ’Visible’, ’On’, ’String’, ’Inf’);
        set(h.lower_cen2, ’Visible’, ’On’);
        set(h.start_cen2, ’Visible’, ’On’);
        set(h.upper_cen2, ’Visible’, ’On’);

        set(h.lower_wid3, ’Visible’, ’Off’);
        set(h.start_wid3, ’Visible’, ’Off’);
        set(h.upper_wid3, ’Visible’, ’Off’);
        set(h.lower_amp3, ’Visible’, ’Off’);
        set(h.start_amp3, ’Visible’, ’Off’);
        set(h.upper_amp3, ’Visible’, ’Off’);
        set(h.lower_cen3, ’Visible’, ’Off’);
        set(h.start_cen3, ’Visible’, ’Off’);
        set(h.upper_cen3, ’Visible’, ’Off’);

        set(h.lower_cen4, ’Visible’, ’Off’);
        set(h.start_cen4, ’Visible’, ’Off’);
        set(h.upper_cen4, ’Visible’, ’Off’);
        set(h.lower_wid4, ’Visible’, ’Off’);
        set(h.start_wid4, ’Visible’, ’Off’);
        set(h.upper_wid4, ’Visible’, ’Off’);
        set(h.lower_amp4, ’Visible’, ’Off’);
        set(h.start_amp4, ’Visible’, ’Off’);
        set(h.upper_amp4, ’Visible’, ’Off’);

        set(h.lower_wid5, ’Visible’, ’Off’);
        set(h.start_wid5, ’Visible’, ’Off’);
        set(h.upper_wid5, ’Visible’, ’Off’);
        set(h.lower_amp5, ’Visible’, ’Off’);
        set(h.start_amp5, ’Visible’, ’Off’);
        set(h.upper_amp5, ’Visible’, ’Off’);
        set(h.lower_cen5, ’Visible’, ’Off’);
        set(h.start_cen5, ’Visible’, ’Off’);
        set(h.upper_cen5, ’Visible’, ’Off’);

        set(h.display_eq, ’String’, ...
            ’y = a1*exp{(x-cen1)^2/wid1} + a2*exp{(x-cen2)^2/wid2}’);
        set(h.amp2, ’String’, ’amplitude 2’, ’TooltipString’, ...
            ’<html><b>Amplitude</b><br/>Defines the amplitude ("height")
                of the Gaussian fit function.<br/>See also "A" in equation (2)
                of the original publication.</html>’);
    end

elseif num_gauss == 3
    tick2 = strcmp(get(h.submenu_kinetics, ’Checked’),’On’);
```

```
if tick2 == 1
     set(h.amp2, 'String', 'amplitude 2', 'TooltipString', ...
          '<html><b>Amplitude</b><br/>Initial occurrence at t = 0.  Adopts a
               value of 1<br/>for normalised cumulative probability
               distributions</html>');
     set(h.display_eq, 'String', ...
          'y = a1*exp{-((x-b1)/tau1)} + ...  + a3*exp{-((x-b3)/tau3)} + c');
     set(h.lower_wid2, 'Visible', 'Off');
     set(h.start_wid2, 'Visible', 'Off');
     set(h.upper_wid2, 'Visible', 'Off');
     set(h.lower_amp2, 'Visible', 'On', 'String', '0.01');
     set(h.start_amp2, 'Visible', 'On', 'String', '0.8');
     set(h.upper_amp2, 'Visible', 'On', 'String', '1');
     set(h.lower_cen2, 'Visible', 'On');
     set(h.start_cen2, 'Visible', 'On');
     set(h.upper_cen2, 'Visible', 'On');

     set(h.lower_wid3, 'Visible', 'Off');
     set(h.start_wid3, 'Visible', 'Off');
     set(h.upper_wid3, 'Visible', 'Off');
     set(h.lower_amp3, 'Visible', 'On');
     set(h.start_amp3, 'Visible', 'On');
     set(h.upper_amp3, 'Visible', 'On');
     set(h.lower_cen3, 'Visible', 'On');
     set(h.start_cen3, 'Visible', 'On');
     set(h.upper_cen3, 'Visible', 'On');

     set(h.lower_cen4, 'Visible', 'Off');
     set(h.start_cen4, 'Visible', 'Off');
     set(h.upper_cen4, 'Visible', 'Off');
     set(h.lower_wid4, 'Visible', 'Off');
     set(h.start_wid4, 'Visible', 'Off');
     set(h.upper_wid4, 'Visible', 'Off');
     set(h.lower_amp4, 'Visible', 'Off');
     set(h.start_amp4, 'Visible', 'Off');
     set(h.upper_amp4, 'Visible', 'Off');

     set(h.lower_wid5, 'Visible', 'Off');
     set(h.start_wid5, 'Visible', 'Off');
     set(h.upper_wid5, 'Visible', 'Off');
     set(h.lower_amp5, 'Visible', 'Off');
     set(h.start_amp5, 'Visible', 'Off');
     set(h.upper_amp5, 'Visible', 'Off');
     set(h.lower_cen5, 'Visible', 'Off');
     set(h.start_cen5, 'Visible', 'Off');
     set(h.upper_cen5, 'Visible', 'Off');

else
     set(h.lower_wid2, 'Visible', 'On');
     set(h.start_wid2, 'Visible', 'On');
```

```
            set(h.upper_wid2, 'Visible', 'On');
            set(h.lower_amp2, 'Visible', 'On', 'String', '0');
            set(h.start_amp2, 'Visible', 'On', 'String', '0.05');
            set(h.upper_amp2, 'Visible', 'On', 'String', 'Inf');
            set(h.lower_cen2, 'Visible', 'On');
            set(h.start_cen2, 'Visible', 'On');
            set(h.upper_cen2, 'Visible', 'On');

            set(h.lower_wid3, 'Visible', 'On');
            set(h.start_wid3, 'Visible', 'On');
            set(h.upper_wid3, 'Visible', 'On');
            set(h.lower_amp3, 'Visible', 'On');
            set(h.start_amp3, 'Visible', 'On');
            set(h.upper_amp3, 'Visible', 'On');
            set(h.lower_cen3, 'Visible', 'On');
            set(h.start_cen3, 'Visible', 'On');
            set(h.upper_cen3, 'Visible', 'On');

            set(h.lower_cen4, 'Visible', 'Off');
            set(h.start_cen4, 'Visible', 'Off');
            set(h.upper_cen4, 'Visible', 'Off');
            set(h.lower_wid4, 'Visible', 'Off');
            set(h.start_wid4, 'Visible', 'Off');
            set(h.upper_wid4, 'Visible', 'Off');
            set(h.lower_amp4, 'Visible', 'Off');
            set(h.start_amp4, 'Visible', 'Off');
            set(h.upper_amp4, 'Visible', 'Off');

            set(h.lower_wid5, 'Visible', 'Off');
            set(h.start_wid5, 'Visible', 'Off');
            set(h.upper_wid5, 'Visible', 'Off');
            set(h.lower_amp5, 'Visible', 'Off');
            set(h.start_amp5, 'Visible', 'Off');
            set(h.upper_amp5, 'Visible', 'Off');
            set(h.lower_cen5, 'Visible', 'Off');
            set(h.start_cen5, 'Visible', 'Off');
            set(h.upper_cen5, 'Visible', 'Off');
            set(h.display_eq, 'String', ...
                'y = a1*exp{(x-cen1)^2/wid1} + ...  + a3*exp{(x-cen3)^2/wid3}');
            set(h.amp2, 'String', 'amplitude 2', 'TooltipString', ...
                '<html><b>Amplitude</b><br/>Defines the amplitude ("height") of
                    the Gaussian fit function.<br/>See also "A" in equation (2)
                    of the original publication.</html>');
        end

elseif num_gauss == 4
        tick2 = strcmp(get(h.submenu_kinetics, 'Checked'),'On');
        if tick2 == 1
            set(h.amp2, 'String', 'amplitude 2', 'TooltipString', ...
                '<html><b>Amplitude</b><br/>Initial occurrence at t = 0.  Adopts a
```

```
                value of 1<br/>for normalised cumulative probability
                distributions</html>');
        set(h.display_eq, 'String', ...
             'y = a1*exp{-((x-b1)/tau1)} + ...  + a4*exp{-((x-b4)/tau4)} + c');

        set(h.lower_wid2, 'Visible', 'Off');
        set(h.start_wid2, 'Visible', 'Off');
        set(h.upper_wid2, 'Visible', 'Off');
        set(h.lower_amp2, 'Visible', 'On', 'String', '0.01');
        set(h.start_amp2, 'Visible', 'On', 'String', '0.8');
        set(h.upper_amp2, 'Visible', 'On', 'String', '1');
        set(h.lower_cen2, 'Visible', 'On');
        set(h.start_cen2, 'Visible', 'On');
        set(h.upper_cen2, 'Visible', 'On');

        set(h.lower_wid3, 'Visible', 'Off');
        set(h.start_wid3, 'Visible', 'Off');
        set(h.upper_wid3, 'Visible', 'Off');
        set(h.lower_amp3, 'Visible', 'On');
        set(h.start_amp3, 'Visible', 'On');
        set(h.upper_amp3, 'Visible', 'On');
        set(h.lower_cen3, 'Visible', 'On');
        set(h.start_cen3, 'Visible', 'On');
        set(h.upper_cen3, 'Visible', 'On');

        set(h.lower_cen4, 'Visible', 'On');
        set(h.start_cen4, 'Visible', 'On');
        set(h.upper_cen4, 'Visible', 'On');
        set(h.lower_wid4, 'Visible', 'Off');
        set(h.start_wid4, 'Visible', 'Off');
        set(h.upper_wid4, 'Visible', 'Off');
        set(h.lower_amp4, 'Visible', 'On');
        set(h.start_amp4, 'Visible', 'On');
        set(h.upper_amp4, 'Visible', 'On');

        set(h.lower_wid5, 'Visible', 'Off');
        set(h.start_wid5, 'Visible', 'Off');
        set(h.upper_wid5, 'Visible', 'Off');
        set(h.lower_amp5, 'Visible', 'Off');
        set(h.start_amp5, 'Visible', 'Off');
        set(h.upper_amp5, 'Visible', 'Off');
        set(h.lower_cen5, 'Visible', 'Off');
        set(h.start_cen5, 'Visible', 'Off');
        set(h.upper_cen5, 'Visible', 'Off');

    else

        set(h.lower_wid2, 'Visible', 'On');
        set(h.start_wid2, 'Visible', 'On');
        set(h.upper_wid2, 'Visible', 'On');
```

```
        set(h.lower_amp2, 'Visible', 'On', 'String', '0');
        set(h.start_amp2, 'Visible', 'On', 'String', '0.05');
        set(h.upper_amp2, 'Visible', 'On', 'String', 'Inf');
        set(h.lower_cen2, 'Visible', 'On');
        set(h.start_cen2, 'Visible', 'On');
        set(h.upper_cen2, 'Visible', 'On');

        set(h.lower_wid3, 'Visible', 'On');
        set(h.start_wid3, 'Visible', 'On');
        set(h.upper_wid3, 'Visible', 'On');
        set(h.lower_amp3, 'Visible', 'On');
        set(h.start_amp3, 'Visible', 'On');
        set(h.upper_amp3, 'Visible', 'On');
        set(h.lower_cen3, 'Visible', 'On');
        set(h.start_cen3, 'Visible', 'On');
        set(h.upper_cen3, 'Visible', 'On');

        set(h.lower_cen4, 'Visible', 'On');
        set(h.start_cen4, 'Visible', 'On');
        set(h.upper_cen4, 'Visible', 'On');
        set(h.lower_wid4, 'Visible', 'On');
        set(h.start_wid4, 'Visible', 'On');
        set(h.upper_wid4, 'Visible', 'On');
        set(h.lower_amp4, 'Visible', 'On');
        set(h.start_amp4, 'Visible', 'On');
        set(h.upper_amp4, 'Visible', 'On');

        set(h.lower_wid5, 'Visible', 'Off');
        set(h.start_wid5, 'Visible', 'Off');
        set(h.upper_wid5, 'Visible', 'Off');
        set(h.lower_amp5, 'Visible', 'Off');
        set(h.start_amp5, 'Visible', 'Off');
        set(h.upper_amp5, 'Visible', 'Off');
        set(h.lower_cen5, 'Visible', 'Off');
        set(h.start_cen5, 'Visible', 'Off');
        set(h.upper_cen5, 'Visible', 'Off');
        set(h.display_eq, 'String', ...
             'y = a1*exp{(x-cen1)^2/wid1} + ...  + a4*exp{(x-cen4)^2/wid4}');
        set(h.amp2, 'String', 'amplitude 2', 'TooltipString', ...
             '<html><b>Amplitude</b><br/>Defines the amplitude ("height") of
                  the Gaussian fit function.<br/>See also "A" in equation (2)
                  of the original publication.</html>');
    end

elseif num_gauss == 5
    tick2 = strcmp(get(h.submenu_kinetics, 'Checked'),'On');
    if tick2 == 1
        set(h.lower_wid2, 'Visible', 'Off');
        set(h.start_wid2, 'Visible', 'Off');
        set(h.upper_wid2, 'Visible', 'Off');
```

```
          set(h.lower_amp2, ’Visible’, ’On’, ’String’, ’0.4’);
          set(h.start_amp2, ’Visible’, ’On’, ’String’, ’0.8’);
          set(h.upper_amp2, ’Visible’, ’On’, ’String’, ’1’);
          set(h.lower_cen2, ’Visible’, ’Off’);
          set(h.start_cen2, ’Visible’, ’Off’);
          set(h.upper_cen2, ’Visible’, ’Off’);

          set(h.lower_wid3, ’Visible’, ’Off’);
          set(h.start_wid3, ’Visible’, ’Off’);
          set(h.upper_wid3, ’Visible’, ’Off’);
          set(h.lower_amp3, ’Visible’, ’Off’);
          set(h.start_amp3, ’Visible’, ’Off’);
          set(h.upper_amp3, ’Visible’, ’Off’);
          set(h.lower_cen3, ’Visible’, ’Off’);
          set(h.start_cen3, ’Visible’, ’Off’);
          set(h.upper_cen3, ’Visible’, ’Off’);

          set(h.lower_cen4, ’Visible’, ’Off’);
          set(h.start_cen4, ’Visible’, ’Off’);
          set(h.upper_cen4, ’Visible’, ’Off’);
          set(h.lower_wid4, ’Visible’, ’Off’);
          set(h.start_wid4, ’Visible’, ’Off’);
          set(h.upper_wid4, ’Visible’, ’Off’);
          set(h.lower_amp4, ’Visible’, ’Off’);
          set(h.start_amp4, ’Visible’, ’Off’);
          set(h.upper_amp4, ’Visible’, ’Off’);

          set(h.lower_wid5, ’Visible’, ’Off’);
          set(h.start_wid5, ’Visible’, ’Off’);
          set(h.upper_wid5, ’Visible’, ’Off’);
          set(h.lower_amp5, ’Visible’, ’Off’);
          set(h.start_amp5, ’Visible’, ’Off’);
          set(h.upper_amp5, ’Visible’, ’Off’);
          set(h.lower_cen5, ’Visible’, ’Off’);
          set(h.start_cen5, ’Visible’, ’Off’);
          set(h.upper_cen5, ’Visible’, ’Off’);

          set(h.amp2, ’String’, ’beta’, ’TooltipString’, ...
               ’<html><b>Amplitude</b><br/>Defines the amplitude ("height") of
                    the Gaussian fit function.<br/>See also "A" in equation (2)
                    of the original publication.</html>’);
          set(h.display_eq, ’String’, ...
               ’y = a1*exp{-((x-b/tau})^beta) + c’);
     else
          set(h.lower_wid2, ’Visible’, ’On’);
          set(h.start_wid2, ’Visible’, ’On’);
          set(h.upper_wid2, ’Visible’, ’On’);
          set(h.lower_amp2, ’Visible’, ’On’, ’String’, ’0’);
          set(h.start_amp2, ’Visible’, ’On’, ’String’, ’0.05’);
          set(h.upper_amp2, ’Visible’, ’On’, ’String’, ’Inf’);
```

```
          set(h.lower_cen2, 'Visible', 'On');
          set(h.start_cen2, 'Visible', 'On');
          set(h.upper_cen2, 'Visible', 'On');

          set(h.lower_wid3, 'Visible', 'On');
          set(h.start_wid3, 'Visible', 'On');
          set(h.upper_wid3, 'Visible', 'On');
          set(h.lower_amp3, 'Visible', 'On');
          set(h.start_amp3, 'Visible', 'On');
          set(h.upper_amp3, 'Visible', 'On');
          set(h.lower_cen3, 'Visible', 'On');
          set(h.start_cen3, 'Visible', 'On');
          set(h.upper_cen3, 'Visible', 'On');

          set(h.lower_cen4, 'Visible', 'On');
          set(h.start_cen4, 'Visible', 'On');
          set(h.upper_cen4, 'Visible', 'On');
          set(h.lower_wid4, 'Visible', 'On');
          set(h.start_wid4, 'Visible', 'On');
          set(h.upper_wid4, 'Visible', 'On');
          set(h.lower_amp4, 'Visible', 'On');
          set(h.start_amp4, 'Visible', 'On');
          set(h.upper_amp4, 'Visible', 'On');

          set(h.lower_wid5, 'Visible', 'On');
          set(h.start_wid5, 'Visible', 'On');
          set(h.upper_wid5, 'Visible', 'On');
          set(h.lower_amp5, 'Visible', 'On');
          set(h.start_amp5, 'Visible', 'On');
          set(h.upper_amp5, 'Visible', 'On');
          set(h.lower_cen5, 'Visible', 'On');
          set(h.start_cen5, 'Visible', 'On');
          set(h.upper_cen5, 'Visible', 'On');
          set(h.display_eq, 'String', ...
               'y = a1*exp{(x-cen1)^2/wid1} + ...  + a5*exp{(x-cen5)^2/wid5}');
          set(h.amp2, 'String', 'amplitude 2', 'TooltipString', ...
               '<html><b>Amplitude</b><br/>Defines the amplitude ("height") of
                    the Gaussian fit function.<br/>See also "A" in equation (2)
                    of the original publication.</html>');
     end
elseif cmpstr(get(h.num_gauss, 'Value'),'single') == 1

end

function num_gauss_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end
```

```
function edit_filetype_Callback(obj, evd, h)
function edit_filetype_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% work in progress
function pushbutton_fitonly_Callback(obj, evd, h)

% Thermodynamic analysis from histogram or DT files?
function radiobutton_histfiles_Callback(obj, evd, h)
radio3 = get(h.radiobutton_histfiles, 'Value');
if radio3 == 1
     set(h.radiobutton_dtfiles, 'Value', 0);
     % show RMSE analysis
     set(h.text29, 'Enable', 'On');
     set(h.text39, 'Enable', 'On');
     set(h.text33, 'Enable', 'On');
     set(h.text32, 'Enable', 'On');
     set(h.text31, 'Enable', 'On');
     set(h.text30, 'Enable', 'On');
     set(h.edit_rms_penalty, 'Enable', 'On');
     set(h.edit_rms_suggestion, 'Enable', 'On');
     set(h.pushbutton_rms, 'Enable', 'On');
     set(h.edit_rms_2gaussians, 'Enable', 'On');
     set(h.edit_rms_3gaussians, 'Enable', 'On');
     set(h.edit_rms_4gaussians, 'Enable', 'On');
     set(h.edit_rms_5gaussians, 'Enable', 'On');
     % show radiobuttons fit type
     set(h.radiobutton_gaussfit, 'Enable', 'On');
     set(h.radiobutton_thresholding, 'Enable', 'On');
     % show edit fields
     set(h.lower_amp1, 'Enable', 'On');
     set(h.start_amp1, 'Enable', 'On');
     set(h.upper_amp1, 'Enable', 'On');
     set(h.lower_cen1, 'Enable', 'On');
     set(h.start_cen1, 'Enable', 'On');
     set(h.upper_cen1, 'Enable', 'On');
     set(h.lower_wid1, 'Enable', 'On');
     set(h.start_wid1, 'Enable', 'On');
     set(h.upper_wid1, 'Enable', 'On');

     set(h.lower_amp2, 'Enable', 'On');
     set(h.start_amp2, 'Enable', 'On');
     set(h.upper_amp2, 'Enable', 'On');
     set(h.lower_cen2, 'Enable', 'On');
     set(h.start_cen2, 'Enable', 'On');
     set(h.upper_cen2, 'Enable', 'On');
     set(h.lower_wid2, 'Enable', 'On');
```

```
     set(h.start_wid2, 'Enable', 'On');
     set(h.upper_wid2, 'Enable', 'On');

     set(h.lower_amp3, 'Enable', 'On');
     set(h.start_amp3, 'Enable', 'On');
     set(h.upper_amp3, 'Enable', 'On');
     set(h.lower_cen3, 'Enable', 'On');
     set(h.start_cen3, 'Enable', 'On');
     set(h.upper_cen3, 'Enable', 'On');
     set(h.lower_wid3, 'Enable', 'On');
     set(h.start_wid3, 'Enable', 'On');
     set(h.upper_wid3, 'Enable', 'On');

     set(h.lower_amp4, 'Enable', 'On');
     set(h.start_amp4, 'Enable', 'On');
     set(h.upper_amp4, 'Enable', 'On');
     set(h.lower_cen4, 'Enable', 'On');
     set(h.start_cen4, 'Enable', 'On');
     set(h.upper_cen4, 'Enable', 'On');
     set(h.lower_wid4, 'Enable', 'On');
     set(h.start_wid4, 'Enable', 'On');
     set(h.upper_wid4, 'Enable', 'On');

     set(h.lower_amp5, 'Enable', 'On');
     set(h.start_amp5, 'Enable', 'On');
     set(h.upper_amp5, 'Enable', 'On');
     set(h.lower_cen5, 'Enable', 'On');
     set(h.start_cen5, 'Enable', 'On');
     set(h.upper_cen5, 'Enable', 'On');
     set(h.lower_wid5, 'Enable', 'On');
     set(h.start_wid5, 'Enable', 'On');
     set(h.upper_wid5, 'Enable', 'On');

     set(h.num_gauss, 'Enable', 'On');
     set(h.display_eq, 'Visible', 'On');
else
     set(h.radiobutton_dtfiles, 'Value', 1);
     % hide RMSE analysis
     set(h.text29, 'Enable', 'Off');
     set(h.text39, 'Enable', 'Off');
     set(h.text33, 'Enable', 'Off');
     set(h.text32, 'Enable', 'Off');
     set(h.text31, 'Enable', 'Off');
     set(h.text30, 'Enable', 'Off');
     set(h.edit_rms_penalty, 'Enable', 'Off');
     set(h.edit_rms_suggestion, 'Enable', 'Off');
     set(h.pushbutton_rms, 'Enable', 'Off');
     set(h.edit_rms_2gaussians, 'Enable', 'Off');
     set(h.edit_rms_3gaussians, 'Enable', 'Off');
     set(h.edit_rms_4gaussians, 'Enable', 'Off');
```

```
set(h.edit_rms_5gaussians, 'Enable', 'Off');
% hide radiobuttons fit type
set(h.radiobutton_gaussfit, 'Enable', 'Off');
set(h.radiobutton_thresholding, 'Enable', 'Off');
% hide edit fields
set(h.lower_amp1, 'Enable', 'Off');
set(h.start_amp1, 'Enable', 'Off');
set(h.upper_amp1, 'Enable', 'Off');
set(h.lower_cen1, 'Enable', 'Off');
set(h.start_cen1, 'Enable', 'Off');
set(h.upper_cen1, 'Enable', 'Off');
set(h.lower_wid1, 'Enable', 'Off');
set(h.start_wid1, 'Enable', 'Off');
set(h.upper_wid1, 'Enable', 'Off');

set(h.lower_amp2, 'Enable', 'Off');
set(h.start_amp2, 'Enable', 'Off');
set(h.upper_amp2, 'Enable', 'Off');
set(h.lower_cen2, 'Enable', 'Off');
set(h.start_cen2, 'Enable', 'Off');
set(h.upper_cen2, 'Enable', 'Off');
set(h.lower_wid2, 'Enable', 'Off');
set(h.start_wid2, 'Enable', 'Off');
set(h.upper_wid2, 'Enable', 'Off');

set(h.lower_amp3, 'Enable', 'Off');
set(h.start_amp3, 'Enable', 'Off');
set(h.upper_amp3, 'Enable', 'Off');
set(h.lower_cen3, 'Enable', 'Off');
set(h.start_cen3, 'Enable', 'Off');
set(h.upper_cen3, 'Enable', 'Off');
set(h.lower_wid3, 'Enable', 'Off');
set(h.start_wid3, 'Enable', 'Off');
set(h.upper_wid3, 'Enable', 'Off');

set(h.lower_amp4, 'Enable', 'Off');
set(h.start_amp4, 'Enable', 'Off');
set(h.upper_amp4, 'Enable', 'Off');
set(h.lower_cen4, 'Enable', 'Off');
set(h.start_cen4, 'Enable', 'Off');
set(h.upper_cen4, 'Enable', 'Off');
set(h.lower_wid4, 'Enable', 'Off');
set(h.start_wid4, 'Enable', 'Off');
set(h.upper_wid4, 'Enable', 'Off');

set(h.lower_amp5, 'Enable', 'Off');
set(h.start_amp5, 'Enable', 'Off');
set(h.upper_amp5, 'Enable', 'Off');
set(h.lower_cen5, 'Enable', 'Off');
set(h.start_cen5, 'Enable', 'Off');
```

```
     set(h.upper_cen5, 'Enable', 'Off');
     set(h.lower_wid5, 'Enable', 'Off');
     set(h.start_wid5, 'Enable', 'Off');
     set(h.upper_wid5, 'Enable', 'Off');

     set(h.num_gauss, 'Enable', 'Off');
     set(h.display_eq, 'Visible', 'Off');
end

function radiobutton_dtfiles_Callback(obj, evd, h)
radio4 = get(h.radiobutton_dtfiles, 'Value');
if radio4 == 1
     set(h.radiobutton_histfiles, 'Value', 0);
     % hide RMSE analysis
     set(h.text29, 'Enable', 'Off');
     set(h.text39, 'Enable', 'Off');
     set(h.text33, 'Enable', 'Off');
     set(h.text32, 'Enable', 'Off');
     set(h.text31, 'Enable', 'Off');
     set(h.text30, 'Enable', 'Off');
     set(h.edit_rms_penalty, 'Enable', 'Off');
     set(h.edit_rms_suggestion, 'Enable', 'Off');
     set(h.pushbutton_rms, 'Enable', 'Off');
     set(h.edit_rms_2gaussians, 'Enable', 'Off');
     set(h.edit_rms_3gaussians, 'Enable', 'Off');
     set(h.edit_rms_4gaussians, 'Enable', 'Off');
     set(h.edit_rms_5gaussians, 'Enable', 'Off');
     % hide radiobuttons fit type
     set(h.radiobutton_gaussfit, 'Enable', 'Off');
     set(h.radiobutton_thresholding, 'Enable', 'Off');
     % hide edit fields
     set(h.lower_amp1, 'Enable', 'Off');
     set(h.start_amp1, 'Enable', 'Off');
     set(h.upper_amp1, 'Enable', 'Off');
     set(h.lower_cen1, 'Enable', 'Off');
     set(h.start_cen1, 'Enable', 'Off');
     set(h.upper_cen1, 'Enable', 'Off');
     set(h.lower_wid1, 'Enable', 'Off');
     set(h.start_wid1, 'Enable', 'Off');
     set(h.upper_wid1, 'Enable', 'Off');

     set(h.lower_amp2, 'Enable', 'Off');
     set(h.start_amp2, 'Enable', 'Off');
     set(h.upper_amp2, 'Enable', 'Off');
     set(h.lower_cen2, 'Enable', 'Off');
     set(h.start_cen2, 'Enable', 'Off');
     set(h.upper_cen2, 'Enable', 'Off');
     set(h.lower_wid2, 'Enable', 'Off');
     set(h.start_wid2, 'Enable', 'Off');
     set(h.upper_wid2, 'Enable', 'Off');
```

```
     set(h.lower_amp3, 'Enable', 'Off');
     set(h.start_amp3, 'Enable', 'Off');
     set(h.upper_amp3, 'Enable', 'Off');
     set(h.lower_cen3, 'Enable', 'Off');
     set(h.start_cen3, 'Enable', 'Off');
     set(h.upper_cen3, 'Enable', 'Off');
     set(h.lower_wid3, 'Enable', 'Off');
     set(h.start_wid3, 'Enable', 'Off');
     set(h.upper_wid3, 'Enable', 'Off');

     set(h.lower_amp4, 'Enable', 'Off');
     set(h.start_amp4, 'Enable', 'Off');
     set(h.upper_amp4, 'Enable', 'Off');
     set(h.lower_cen4, 'Enable', 'Off');
     set(h.start_cen4, 'Enable', 'Off');
     set(h.upper_cen4, 'Enable', 'Off');
     set(h.lower_wid4, 'Enable', 'Off');
     set(h.start_wid4, 'Enable', 'Off');
     set(h.upper_wid4, 'Enable', 'Off');

     set(h.lower_amp5, 'Enable', 'Off');
     set(h.start_amp5, 'Enable', 'Off');
     set(h.upper_amp5, 'Enable', 'Off');
     set(h.lower_cen5, 'Enable', 'Off');
     set(h.start_cen5, 'Enable', 'Off');
     set(h.upper_cen5, 'Enable', 'Off');
     set(h.lower_wid5, 'Enable', 'Off');
     set(h.start_wid5, 'Enable', 'Off');
     set(h.upper_wid5, 'Enable', 'Off');

     set(h.num_gauss, 'Enable', 'Off');
     set(h.display_eq, 'Visible', 'Off');
else
     set(h.radiobutton_histfiles, 'Value', 1);
     % show RMSE analysis
     set(h.text29, 'Enable', 'On');
     set(h.text39, 'Enable', 'On');
     set(h.text33, 'Enable', 'On');
     set(h.text32, 'Enable', 'On');
     set(h.text31, 'Enable', 'On');
     set(h.text30, 'Enable', 'On');
     set(h.edit_rms_penalty, 'Enable', 'On');
     set(h.edit_rms_suggestion, 'Enable', 'On');
     set(h.pushbutton_rms, 'Enable', 'On');
     set(h.edit_rms_2gaussians, 'Enable', 'On');
     set(h.edit_rms_3gaussians, 'Enable', 'On');
     set(h.edit_rms_4gaussians, 'Enable', 'On');
     set(h.edit_rms_5gaussians, 'Enable', 'On');
     % show radiobuttons fit type
```

```
set(h.radiobutton_gaussfit, 'Enable', 'On');
set(h.radiobutton_thresholding, 'Enable', 'On');
% show edit fields
set(h.lower_amp1, 'Enable', 'On');
set(h.start_amp1, 'Enable', 'On');
set(h.upper_amp1, 'Enable', 'On');
set(h.lower_cen1, 'Enable', 'On');
set(h.start_cen1, 'Enable', 'On');
set(h.upper_cen1, 'Enable', 'On');
set(h.lower_wid1, 'Enable', 'On');
set(h.start_wid1, 'Enable', 'On');
set(h.upper_wid1, 'Enable', 'On');

set(h.lower_amp2, 'Enable', 'On');
set(h.start_amp2, 'Enable', 'On');
set(h.upper_amp2, 'Enable', 'On');
set(h.lower_cen2, 'Enable', 'On');
set(h.start_cen2, 'Enable', 'On');
set(h.upper_cen2, 'Enable', 'On');
set(h.lower_wid2, 'Enable', 'On');
set(h.start_wid2, 'Enable', 'On');
set(h.upper_wid2, 'Enable', 'On');

set(h.lower_amp3, 'Enable', 'On');
set(h.start_amp3, 'Enable', 'On');
set(h.upper_amp3, 'Enable', 'On');
set(h.lower_cen3, 'Enable', 'On');
set(h.start_cen3, 'Enable', 'On');
set(h.upper_cen3, 'Enable', 'On');
set(h.lower_wid3, 'Enable', 'On');
set(h.start_wid3, 'Enable', 'On');
set(h.upper_wid3, 'Enable', 'On');

set(h.lower_amp4, 'Enable', 'On');
set(h.start_amp4, 'Enable', 'On');
set(h.upper_amp4, 'Enable', 'On');
set(h.lower_cen4, 'Enable', 'On');
set(h.start_cen4, 'Enable', 'On');
set(h.upper_cen4, 'Enable', 'On');
set(h.lower_wid4, 'Enable', 'On');
set(h.start_wid4, 'Enable', 'On');
set(h.upper_wid4, 'Enable', 'On');

set(h.lower_amp5, 'Enable', 'On');
set(h.start_amp5, 'Enable', 'On');
set(h.upper_amp5, 'Enable', 'On');
set(h.lower_cen5, 'Enable', 'On');
set(h.start_cen5, 'Enable', 'On');
set(h.upper_cen5, 'Enable', 'On');
set(h.lower_wid5, 'Enable', 'On');
```

```
     set(h.start_wid5, 'Enable', 'On');
     set(h.upper_wid5, 'Enable', 'On');

     set(h.num_gauss, 'Enable', 'On');
     set(h.display_eq, 'Visible', 'On');
end

% weighing yes/no
function radiobutton_weighing_yes_Callback(obj, evd, h)
radio1 = get(h.radiobutton_weighing_yes, 'Value');
if radio1 == 1
     set(h.radiobutton_weighing_no, 'Value', 0);
else
     set(h.radiobutton_weighing_no, 'Value', 1);
end

function radiobutton_weighing_no_Callback(obj, evd, h)
radio2 = get(h.radiobutton_weighing_yes, 'Value');
if radio2 == 1
     set(h.radiobutton_weighing_yes, 'Value', 0);
else
     set(h.radiobutton_weighing_yes, 'Value', 1);
end

% Gaussian fitting or thresholding?
function radiobutton_gaussfit_Callback(obj, evd, h)
radio5 = get(h.radiobutton_gaussfit, 'Value');
if radio5 == 1
     set(h.radiobutton_thresholding, 'Value', 0);
     set(h.num_gauss, 'String', {'number of Gaussians';'2';'3'; ...
          '4';'5'}, 'Value', 1);
     % show RMSE analysis
     set(h.text29, 'Enable', 'On');
     set(h.text39, 'Enable', 'On');
     set(h.text33, 'Enable', 'On');
     set(h.text32, 'Enable', 'On');
     set(h.text31, 'Enable', 'On');
     set(h.text30, 'Enable', 'On');
     set(h.edit_rms_penalty, 'Enable', 'On');
     set(h.edit_rms_suggestion, 'Enable', 'On');
     set(h.pushbutton_rms, 'Enable', 'On');
     set(h.edit_rms_2gaussians, 'Enable', 'On');
     set(h.edit_rms_3gaussians, 'Enable', 'On');
     set(h.edit_rms_4gaussians, 'Enable', 'On');
     set(h.edit_rms_5gaussians, 'Enable', 'On');
     % show edit fields
     set(h.lower_amp1, 'Enable', 'On');
     set(h.start_amp1, 'Enable', 'On');
     set(h.upper_amp1, 'Enable', 'On');
     set(h.lower_cen1, 'Enable', 'On');
```

```
     set(h.start_cen1, ’Enable’, ’On’);
     set(h.upper_cen1, ’Enable’, ’On’);
     set(h.lower_wid1, ’Enable’, ’On’);
     set(h.start_wid1, ’Enable’, ’On’);
     set(h.upper_wid1, ’Enable’, ’On’);

     set(h.lower_amp2, ’Enable’, ’On’);
     set(h.start_amp2, ’Enable’, ’On’);
     set(h.upper_amp2, ’Enable’, ’On’);
     set(h.lower_cen2, ’Enable’, ’On’);
     set(h.start_cen2, ’Enable’, ’On’);
     set(h.upper_cen2, ’Enable’, ’On’);
     set(h.lower_wid2, ’Enable’, ’On’);
     set(h.start_wid2, ’Enable’, ’On’);
     set(h.upper_wid2, ’Enable’, ’On’);

     set(h.lower_amp3, ’Enable’, ’On’);
     set(h.start_amp3, ’Enable’, ’On’);
     set(h.upper_amp3, ’Enable’, ’On’);
     set(h.lower_cen3, ’Enable’, ’On’);
     set(h.start_cen3, ’Enable’, ’On’);
     set(h.upper_cen3, ’Enable’, ’On’);
     set(h.lower_wid3, ’Enable’, ’On’);
     set(h.start_wid3, ’Enable’, ’On’);
     set(h.upper_wid3, ’Enable’, ’On’);

     set(h.lower_amp4, ’Enable’, ’On’);
     set(h.start_amp4, ’Enable’, ’On’);
     set(h.upper_amp4, ’Enable’, ’On’);
     set(h.lower_cen4, ’Enable’, ’On’);
     set(h.start_cen4, ’Enable’, ’On’);
     set(h.upper_cen4, ’Enable’, ’On’);
     set(h.lower_wid4, ’Enable’, ’On’);
     set(h.start_wid4, ’Enable’, ’On’);
     set(h.upper_wid4, ’Enable’, ’On’);

     set(h.lower_amp5, ’Enable’, ’On’);
     set(h.start_amp5, ’Enable’, ’On’);
     set(h.upper_amp5, ’Enable’, ’On’);
     set(h.lower_cen5, ’Enable’, ’On’);
     set(h.start_cen5, ’Enable’, ’On’);
     set(h.upper_cen5, ’Enable’, ’On’);
     set(h.lower_wid5, ’Enable’, ’On’);
     set(h.start_wid5, ’Enable’, ’On’);
     set(h.upper_wid5, ’Enable’, ’On’);

     set(h.display_eq, ’Visible’, ’On’, ’String’, ”);
else
     set(h.radiobutton_thresholding, ’Value’, 1);
     % update dropdown menu
```

```
set(h.num_gauss, 'String', {'1 threshold';'2 thresholds'; ...
     '3 thresholds';'4 thresholds';'5 thresholds'}, 'Value', 1);
% hide RMSE analysis
set(h.text29, 'Enable', 'Off');
set(h.text39, 'Enable', 'Off');
set(h.text33, 'Enable', 'Off');
set(h.text32, 'Enable', 'Off');
set(h.text31, 'Enable', 'Off');
set(h.text30, 'Enable', 'Off');
set(h.edit_rms_penalty, 'Enable', 'Off');
set(h.edit_rms_suggestion, 'Enable', 'Off');
set(h.pushbutton_rms, 'Enable', 'Off');
set(h.edit_rms_2gaussians, 'Enable', 'Off');
set(h.edit_rms_3gaussians, 'Enable', 'Off');
set(h.edit_rms_4gaussians, 'Enable', 'Off');
set(h.edit_rms_5gaussians, 'Enable', 'Off');
% hide edit fields
set(h.lower_amp1, 'Enable', 'Off');
set(h.start_amp1, 'Enable', 'Off');
set(h.upper_amp1, 'Enable', 'Off');
set(h.lower_cen1, 'Enable', 'Off');
set(h.start_cen1, 'Enable', 'Off');
set(h.upper_cen1, 'Enable', 'Off');
set(h.lower_wid1, 'Enable', 'Off');
set(h.start_wid1, 'Enable', 'Off');
set(h.upper_wid1, 'Enable', 'Off');

set(h.lower_amp2, 'Enable', 'Off');
set(h.start_amp2, 'Enable', 'Off');
set(h.upper_amp2, 'Enable', 'Off');
set(h.lower_cen2, 'Enable', 'Off');
set(h.start_cen2, 'Enable', 'Off');
set(h.upper_cen2, 'Enable', 'Off');
set(h.lower_wid2, 'Enable', 'Off');
set(h.start_wid2, 'Enable', 'Off');
set(h.upper_wid2, 'Enable', 'Off');

set(h.lower_amp3, 'Enable', 'Off');
set(h.start_amp3, 'Enable', 'Off');
set(h.upper_amp3, 'Enable', 'Off');
set(h.lower_cen3, 'Enable', 'Off');
set(h.start_cen3, 'Enable', 'Off');
set(h.upper_cen3, 'Enable', 'Off');
set(h.lower_wid3, 'Enable', 'Off');
set(h.start_wid3, 'Enable', 'Off');
set(h.upper_wid3, 'Enable', 'Off');

set(h.lower_amp4, 'Enable', 'Off');
set(h.start_amp4, 'Enable', 'Off');
set(h.upper_amp4, 'Enable', 'Off');
```

```
     set(h.lower_cen4, 'Enable', 'Off');
     set(h.start_cen4, 'Enable', 'Off');
     set(h.upper_cen4, 'Enable', 'Off');
     set(h.lower_wid4, 'Enable', 'Off');
     set(h.start_wid4, 'Enable', 'Off');
     set(h.upper_wid4, 'Enable', 'Off');

     set(h.lower_amp5, 'Enable', 'Off');
     set(h.start_amp5, 'Enable', 'Off');
     set(h.upper_amp5, 'Enable', 'Off');
     set(h.lower_cen5, 'Enable', 'Off');
     set(h.start_cen5, 'Enable', 'Off');
     set(h.upper_cen5, 'Enable', 'Off');
     set(h.lower_wid5, 'Enable', 'Off');
     set(h.start_wid5, 'Enable', 'Off');
     set(h.upper_wid5, 'Enable', 'Off');

     set(h.display_eq, 'Visible', 'Off');
end

function radiobutton_thresholding_Callback(obj, evd, h)
radio6 = get(h.radiobutton_thresholding, 'Value');
if radio6 == 1
     set(h.radiobutton_gaussfit, 'Value', 0);
     set(h.num_gauss, 'String', {'1 threshold';'2 thresholds'; ...
          '3 thresholds';'4 thresholds';'5 thresholds'}, 'Value', 1);
     % hide RMSE analysis
     set(h.text29, 'Enable', 'Off');
     set(h.text39, 'Enable', 'Off');
     set(h.text33, 'Enable', 'Off');
     set(h.text32, 'Enable', 'Off');
     set(h.text31, 'Enable', 'Off');
     set(h.text30, 'Enable', 'Off');
     set(h.edit_rms_penalty, 'Enable', 'Off');
     set(h.edit_rms_suggestion, 'Enable', 'Off');
     set(h.pushbutton_rms, 'Enable', 'Off');
     set(h.edit_rms_2gaussians, 'Enable', 'Off');
     set(h.edit_rms_3gaussians, 'Enable', 'Off');
     set(h.edit_rms_4gaussians, 'Enable', 'Off');
     set(h.edit_rms_5gaussians, 'Enable', 'Off');
     % hide edit fields
     set(h.lower_amp1, 'Enable', 'Off');
     set(h.start_amp1, 'Enable', 'Off');
     set(h.upper_amp1, 'Enable', 'Off');
     set(h.lower_cen1, 'Enable', 'Off');
     set(h.start_cen1, 'Enable', 'Off');
     set(h.upper_cen1, 'Enable', 'Off');
     set(h.lower_wid1, 'Enable', 'Off');
     set(h.start_wid1, 'Enable', 'Off');
     set(h.upper_wid1, 'Enable', 'Off');
```

```
     set(h.lower_amp2, 'Enable', 'Off');
     set(h.start_amp2, 'Enable', 'Off');
     set(h.upper_amp2, 'Enable', 'Off');
     set(h.lower_cen2, 'Enable', 'Off');
     set(h.start_cen2, 'Enable', 'Off');
     set(h.upper_cen2, 'Enable', 'Off');
     set(h.lower_wid2, 'Enable', 'Off');
     set(h.start_wid2, 'Enable', 'Off');
     set(h.upper_wid2, 'Enable', 'Off');

     set(h.lower_amp3, 'Enable', 'Off');
     set(h.start_amp3, 'Enable', 'Off');
     set(h.upper_amp3, 'Enable', 'Off');
     set(h.lower_cen3, 'Enable', 'Off');
     set(h.start_cen3, 'Enable', 'Off');
     set(h.upper_cen3, 'Enable', 'Off');
     set(h.lower_wid3, 'Enable', 'Off');
     set(h.start_wid3, 'Enable', 'Off');
     set(h.upper_wid3, 'Enable', 'Off');

     set(h.lower_amp4, 'Enable', 'Off');
     set(h.start_amp4, 'Enable', 'Off');
     set(h.upper_amp4, 'Enable', 'Off');
     set(h.lower_cen4, 'Enable', 'Off');
     set(h.start_cen4, 'Enable', 'Off');
     set(h.upper_cen4, 'Enable', 'Off');
     set(h.lower_wid4, 'Enable', 'Off');
     set(h.start_wid4, 'Enable', 'Off');
     set(h.upper_wid4, 'Enable', 'Off');

     set(h.lower_amp5, 'Enable', 'Off');
     set(h.start_amp5, 'Enable', 'Off');
     set(h.upper_amp5, 'Enable', 'Off');
     set(h.lower_cen5, 'Enable', 'Off');
     set(h.start_cen5, 'Enable', 'Off');
     set(h.upper_cen5, 'Enable', 'Off');
     set(h.lower_wid5, 'Enable', 'Off');
     set(h.start_wid5, 'Enable', 'Off');
     set(h.upper_wid5, 'Enable', 'Off');

     set(h.display_eq, 'Visible', 'Off');
else
     set(h.radiobutton_gaussfit, 'Value', 1);
     set(h.num_gauss, 'String', {'number of Gaussians';'2';'3'; ...
          '4';'5'}, 'Value', 1);
     % show RMSE analysis
     set(h.text29, 'Enable', 'On');
     set(h.text39, 'Enable', 'On');
     set(h.text33, 'Enable', 'On');
```

```
    set(h.text32, 'Enable', 'On');
    set(h.text31, 'Enable', 'On');
    set(h.text30, 'Enable', 'On');
    set(h.edit_rms_penalty, 'Enable', 'On');
    set(h.edit_rms_suggestion, 'Enable', 'On');
    set(h.pushbutton_rms, 'Enable', 'On');
    set(h.edit_rms_2gaussians, 'Enable', 'On');
    set(h.edit_rms_3gaussians, 'Enable', 'On');
    set(h.edit_rms_4gaussians, 'Enable', 'On');
    set(h.edit_rms_5gaussians, 'Enable', 'On');
    % show edit fields
    set(h.lower_amp1, 'Enable', 'On');
    set(h.start_amp1, 'Enable', 'On');
    set(h.upper_amp1, 'Enable', 'On');
    set(h.lower_cen1, 'Enable', 'On');
    set(h.start_cen1, 'Enable', 'On');
    set(h.upper_cen1, 'Enable', 'On');
    set(h.lower_wid1, 'Enable', 'On');
    set(h.start_wid1, 'Enable', 'On');
    set(h.upper_wid1, 'Enable', 'On');

    set(h.lower_amp2, 'Enable', 'On');
    set(h.start_amp2, 'Enable', 'On');
    set(h.upper_amp2, 'Enable', 'On');
    set(h.lower_cen2, 'Enable', 'On');
    set(h.start_cen2, 'Enable', 'On');
    set(h.upper_cen2, 'Enable', 'On');
    set(h.lower_wid2, 'Enable', 'On');
    set(h.start_wid2, 'Enable', 'On');
    set(h.upper_wid2, 'Enable', 'On');

    set(h.lower_amp3, 'Enable', 'On');
    set(h.start_amp3, 'Enable', 'On');
    set(h.upper_amp3, 'Enable', 'On');
    set(h.lower_cen3, 'Enable', 'On');
    set(h.start_cen3, 'Enable', 'On');
    set(h.upper_cen3, 'Enable', 'On');
    set(h.lower_wid3, 'Enable', 'On');
    set(h.start_wid3, 'Enable', 'On');
    set(h.upper_wid3, 'Enable', 'On');

    set(h.lower_amp4, 'Enable', 'On');
    set(h.start_amp4, 'Enable', 'On');
    set(h.upper_amp4, 'Enable', 'On');
    set(h.lower_cen4, 'Enable', 'On');
    set(h.start_cen4, 'Enable', 'On');
    set(h.upper_cen4, 'Enable', 'On');
    set(h.lower_wid4, 'Enable', 'On');
    set(h.start_wid4, 'Enable', 'On');
    set(h.upper_wid4, 'Enable', 'On');
```

```
     set(h.lower_amp5, ’Enable’, ’On’);
     set(h.start_amp5, ’Enable’, ’On’);
     set(h.upper_amp5, ’Enable’, ’On’);
     set(h.lower_cen5, ’Enable’, ’On’);
     set(h.start_cen5, ’Enable’, ’On’);
     set(h.upper_cen5, ’Enable’, ’On’);
     set(h.lower_wid5, ’Enable’, ’On’);
     set(h.start_wid5, ’Enable’, ’On’);
     set(h.upper_wid5, ’Enable’, ’On’);

     set(h.display_eq, ’Visible’, ’On’, ’String’, ”);
end

% display histogram files in root folder
function listbox_filelist_Callback(obj, evd, h)
function listbox_filelist_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,’BackgroundColor’), ...
          get(0,’defaultUicontrolBackgroundColor’))
     set(obj,’BackgroundColor’,’white’);
end


% number of histogram files
function edit_numfiles2_Callback(obj, evd, h)
function edit_numfiles2_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,’BackgroundColor’), ...
          get(0,’defaultUicontrolBackgroundColor’))
     set(obj,’BackgroundColor’,’white’);
end


% RMSE analysis
function edit_rms_suggestion_Callback(obj, evd, h)
function edit_rms_suggestion_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,’BackgroundColor’), ...
          get(0,’defaultUicontrolBackgroundColor’))
     set(obj,’BackgroundColor’,’white’);
end

function edit_rms_penalty_Callback(obj, evd, h)
function edit_rms_penalty_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,’BackgroundColor’), ...
          get(0,’defaultUicontrolBackgroundColor’))
     set(obj,’BackgroundColor’,’white’);
end

function edit_rms_5gaussians_Callback(obj, evd, h)
function edit_rms_5gaussians_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,’BackgroundColor’), ...
```

```
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

function edit_rms_4gaussians_Callback(obj, evd, h)
function edit_rms_4gaussians_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

function edit_rms_3gaussians_Callback(obj, evd, h)
function edit_rms_3gaussians_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

function edit_rms_2gaussians_Callback(obj, evd, h)
function edit_rms_2gaussians_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end


% About Boba FRET
function menu_file_Callback(obj, evd, h)

function menu_help_Callback(obj, evd, h)

function submenu_about_Callback(obj, evd, h)
about = sprintf('Boba FRET was written by Sebastian König', ...
     'Erica Fiorini, Melodie Hadzic, Danny Kowerko and Richard Börner.');
uiwait(helpdlg(about));
return;

% Switching between Gaussian and exponential fitting
% exec.  of the the thermodyn.  subprogramme
function submenu_thermodynamics_Callback(obj, evd, h)
tick1 = get(h.submenu_thermodynamics, 'Checked');
if strcmp(tick1,'On') == 1
     set(h.submenu_kinetics, 'Checked', 'Off');
else
     % only one item can be ticked at a time
     set(h.submenu_thermodynamics, 'Checked', 'On');
     set(h.submenu_kinetics, 'Checked', 'Off');
     % data import radiobuttons
     set(h.radiobutton_histfiles, 'Enable', 'On', 'Value', 1);
     set(h.radiobutton_dtfiles, 'Enable', 'On', 'Value', 0);
```

```
set(h.radiobutton_weighing_no, 'Value', 0);
set(h.radiobutton_weighing_yes, 'Value', 1);
% visibility of the descriptions
set(h.amp1, 'Visible', 'On', 'String', 'amplitude 1', 'TooltipString',...
     '<html><b>Amplitude</b><br/>Defines the amplitude ("height") of the
     Gaussian fit function.<br/>See also "A" in equation (2) of the
     original publication.</html>');
set(h.amp2, 'Visible', 'On', 'String', 'amplitude 2', 'TooltipString',...
     '<html><b>Amplitude</b><br/>Defines the amplitude ("height") of the
     Gaussian fit function.<br/>See also "A" in equation (2) of the
     original publication.</html>');
set(h.amp3, 'Visible', 'On', 'String', 'amplitude 3', 'TooltipString',...
     '<html><b>Amplitude</b><br/>Defines the amplitude ("height") of the
     Gaussian fit function.<br/>See also "A" in equation (2) of the
     original publication.</html>');
set(h.amp4, 'Visible', 'On', 'String', 'amplitude 4', 'TooltipString',...
     '<html><b>Amplitude</b><br/>Defines the amplitude ("height") of the
     Gaussian fit function.<br/>See also "A" in equation (2) of the
     original publication.</html>');
set(h.amp5, 'Visible', 'On', 'String', 'amplitude 5', 'TooltipString',...
     '<html><b>Amplitude</b><br/>Defines the amplitude ("height") of the
     Gaussian fit function.<br/>See also "A" in equation (2) of the
     original publication.</html>', ...
     'Position', [1.0 5.38 11.2 1.0769]);
set(h.cen1, 'Visible', 'On', 'String', 'center 1', 'TooltipString', ...
     '<html><b>Centre</b><br/>Defines the centre ("FRET value") of the
     Gaussian fit function.<br/>See also "b" in equation (2) of the
     original publication.</html>');
set(h.cen2, 'Visible', 'On', 'String', 'center 2', 'TooltipString', ...
     '<html><b>Centre</b><br/>Defines the centre ("FRET value") of the
     Gaussian fit function.<br/>See also "b" in equation (2) of the
     original publication.</html>');
set(h.cen3, 'Visible', 'On', 'String', 'center 3', 'TooltipString', ...
     '<html><b>Centre</b><br/>Defines the centre ("FRET value") of the
     Gaussian fit function.<br/>See also "b" in equation (2) of the
     original publication.</html>');
set(h.cen4, 'Visible', 'On', 'String', 'center 4', 'TooltipString', ...
     '<html><b>Centre</b><br/>Defines the centre ("FRET value") of the
     Gaussian fit function.<br/>See also "b" in equation (2) of the
     original publication.</html>');
set(h.cen5, 'Visible', 'On', 'String', 'center 5');
set(h.wid1, 'Visible', 'On', 'String', 'width 1', 'TooltipString', ...
     '<html><b>Width</b><br/>Defines the standard deviation ("width") of
     the Gaussian fit function.<br/>See also "sigma" in equation (2) of the
     original publication.</html>');
set(h.wid2, 'Visible', 'On', 'String', 'width 2', 'TooltipString', ...
     '<html><b>Width</b><br/>Defines the standard deviation ("width") of
     the Gaussian fit function.<br/>See also "sigma" in equation (2) of the
     original publication.</html>');
set(h.wid3, 'Visible', 'On', 'String', 'width 3', 'TooltipString', ...
```

```
          '<html><b>Width</b><br/>Defines the standard deviation ("width") of
          the Gaussian fit function.<br/>See also "sigma" in equation (2) of the
          original publication.</html>');
     set(h.wid4, 'Visible', 'On', 'String', 'width 4', 'TooltipString', ...
          '<html><b>Width</b><br/>Defines the standard deviation ("width") of
          the Gaussian fit function.<br/>See also "sigma" in equation (2) of the
          original publication.</html>', ...
          'Position', [1.0 7.30769 11.2 1.0769]);
     set(h.wid5, 'Visible', 'On', 'String', 'width 5');
     % offset checkbox and boundaries
     set(h.y_offset, 'Visible', 'Off');
     set(h.edit_lower_y_offset, 'Visible', 'Off');
     set(h.edit_start_y_offset, 'Visible', 'Off');
     set(h.edit_upper_y_offset, 'Visible', 'Off');
     set(h.x_offset, 'Visible', 'Off');
     set(h.edit_lower_x_offset, 'Visible', 'Off');
     set(h.edit_start_x_offset, 'Visible', 'Off');
     set(h.edit_upper_x_offset, 'Visible', 'Off');
     % adjust visibility of text fields
     set(h.lower_amp1, 'Visible', 'On', 'String', '0', ...
          'Enable', 'On');
     set(h.lower_amp2, 'Visible', 'On', 'String', '0', ...
          'Enable', 'On');
     set(h.lower_amp3, 'Visible', 'Off', 'String', '0', ...
          'Enable', 'On');
     set(h.lower_amp4, 'Visible', 'Off', 'String', '0', ...
          'Enable', 'On');
     set(h.lower_amp5, 'Visible', 'Off', 'String', '0', ...
          'Enable', 'On');
     set(h.start_amp1, 'Visible', 'On', 'String', '0.05', ...
          'Enable', 'On');
     set(h.start_amp2, 'Visible', 'On', 'String', '0.05', ...
          'Enable', 'On');
     set(h.start_amp3, 'Visible', 'Off', 'String', '0.05', ...
          'Enable', 'On');
     set(h.start_amp4, 'Visible', 'Off', 'String', '0.05', ...
          'Enable', 'On');
     set(h.start_amp5, 'Visible', 'Off', 'String', '0.05', ...
          'Enable', 'On');
     set(h.upper_amp1, 'Visible', 'On', 'String', 'Inf', ...
          'Enable', 'On');
     set(h.upper_amp2, 'Visible', 'On', 'String', 'Inf', ...
          'Enable', 'On');
     set(h.upper_amp3, 'Visible', 'Off', 'String','Inf', ...
          'Enable', 'On');
     set(h.upper_amp4, 'Visible', 'Off', 'String', 'Inf', ...
          'Enable', 'On');
     set(h.upper_amp5, 'Visible', 'Off', 'String', 'Inf', ...
          'Enable', 'On');
     set(h.lower_cen1, 'Visible', 'On', 'String', '-Inf', ...
```

```
     'Enable', 'On');
set(h.lower_cen2, 'Visible', 'On', 'String', '-Inf', ...
     'Enable', 'On');
set(h.lower_cen3, 'Visible', 'Off', 'String', '-Inf', ...
     'Enable', 'On');
set(h.lower_cen4, 'Visible', 'Off', 'String', '-Inf', ...
     'Enable', 'On');
set(h.lower_cen5, 'Visible', 'Off', 'String', '-Inf', ...
     'Enable', 'On');
set(h.start_cen1, 'Visible', 'On', 'String', '0', ...
     'Enable', 'On');
set(h.start_cen2, 'Visible', 'On', 'String', '0.2', ...
     'Enable', 'On');
set(h.start_cen3, 'Visible', 'Off', 'String', '0.4', ...
     'Enable', 'On');
set(h.start_cen4, 'Visible', 'Off', 'String', '0.6', ...
     'Enable', 'On');
set(h.start_cen5, 'Visible', 'Off', 'String', '0.8', ...
     'Enable', 'On');
set(h.upper_cen1, 'Visible', 'On', 'String', 'Inf', ...
     'Enable', 'On');
set(h.upper_cen2, 'Visible', 'On', 'String', 'Inf', ...
     'Enable', 'On');
set(h.upper_cen3, 'Visible', 'Off', 'String', 'Inf', ...
     'Enable', 'On');
set(h.upper_cen4, 'Visible', 'Off', 'String', 'Inf', ...
     'Enable', 'On');
set(h.upper_cen5, 'Visible', 'Off', 'String', 'Inf', ...
     'Enable', 'On');
set(h.lower_wid1, 'Visible', 'On', 'String', '0', ...
     'Enable', 'On');
set(h.lower_wid2, 'Visible', 'On', 'String', '0', ...
     'Enable', 'On');
set(h.lower_wid3, 'Visible', 'Off', 'String', '0', ...
     'Enable', 'On');
set(h.lower_wid4, 'Visible', 'Off', 'String', '0', ...
     'Enable', 'On');
set(h.lower_wid5, 'Visible', 'Off', 'String', '0', ...
     'Enable', 'On');
set(h.start_wid1, 'Visible', 'On', 'String', '0.087', ...
     'Enable', 'On');
set(h.start_wid2, 'Visible', 'On', 'String', '0.087', ...
     'Enable', 'On');
set(h.start_wid3, 'Visible', 'Off', 'String', '0.087', ...
     'Enable', 'On');
set(h.start_wid4, 'Visible', 'Off', 'String', '0.087', ...
     'Enable', 'On');
set(h.start_wid5, 'Visible', 'Off', 'String', '0.087', ...
     'Enable', 'On');
set(h.upper_wid1, 'Visible', 'On', 'String', '0.15', ...
```

```
          'Enable', 'On');
     set(h.upper_wid2, 'Visible', 'On', 'String', '0.15', ...
          'Enable', 'On');
     set(h.upper_wid3, 'Visible', 'Off', 'String', '0.15', ...
          'Enable', 'On');
     set(h.upper_wid4, 'Visible', 'Off', 'String', '0.15', ...
          'Enable', 'On');
     set(h.upper_wid5, 'Visible', 'Off', 'String', '0.15', ...
          'Enable', 'On');
     % adjust dropdown menu
     set(h.num_gauss, 'String', {'number of Gaussians';'2';'3';'4';'5'}, 'Value', 1);
     % hide equations
     set(h.display_eq, 'Visible', 'On');
     % show radiobuttons Gaussian fitting/thresholding
     set(h.radiobutton_gaussfit, 'Visible', 'On', 'Value', 1, 'Enable', 'On');
     set(h.radiobutton_thresholding, 'Visible', 'On', 'Value', 0, 'Enable', 'On');
     % hide equations
     set(h.display_eq, 'Visible', 'On', 'String', ");
     % show RMSE analysis
     set(h.text29, 'Enable', 'On');
     set(h.text39, 'Enable', 'On');
     set(h.text33, 'Enable', 'On');
     set(h.text32, 'Enable', 'On');
     set(h.text31, 'Enable', 'On');
     set(h.text30, 'Enable', 'On');
     set(h.edit_rms_penalty, 'Enable', 'On');
     set(h.edit_rms_suggestion, 'Enable', 'On');
     set(h.pushbutton_rms, 'Enable', 'On');
     set(h.edit_rms_2gaussians, 'Enable', 'On');
     set(h.edit_rms_3gaussians, 'Enable', 'On');
     set(h.edit_rms_4gaussians, 'Enable', 'On');
     set(h.edit_rms_5gaussians, 'Enable', 'On');
end


% execution of the the kinetics subprogramme
function submenu_kinetics_Callback(obj, evd, h)
tick2 = get(h.submenu_kinetics, 'Checked');
if strcmp(tick2,'On') == 1
     set(h.submenu_thermodynamics, 'Checked', 'Off');
else
     set(h.num_gauss, 'Enable', 'On');
     % only one item can be ticked at a time
     set(h.submenu_kinetics, 'Checked', 'On');
     set(h.submenu_thermodynamics, 'Checked', 'Off');
     % data import radiobuttons
     set(h.radiobutton_histfiles, 'Enable', 'Off');
     set(h.radiobutton_dtfiles, 'Enable', 'Off');
     % visibility of the descriptions
     set(h.amp1, 'Visible', 'On', 'String', 'amplitude 1', 'TooltipString', ...
```

```
        '<html><b>Amplitude</b><br/>Initial occurrence at t = 0.  Adopts a
        value of 1<br/>for normalised cumulative probability
        distributions</html>');
    set(h.amp2, 'Visible', 'On', 'String', 'amplitude 2', 'TooltipString', ...
        '<html><b>Amplitude</b><br/>Initial occurrence at t = 0.  Adopts a
        value of 1<br/>for normalised cumulative probability
        distributions</html>');
    set(h.amp3, 'Visible', 'On', 'String', 'amplitude 3', 'TooltipString', ...
        '<html><b>Amplitude</b><br/>Initial occurrence at t = 0.  Adopts a
        value of 1<br/>for normalised cumulative probability
        distributions</html>');
    set(h.amp4, 'Visible', 'On', 'String', 'amplitude 4', 'TooltipString', ...
        '<html><b>Amplitude</b><br/>Initial occurrence at t = 0.  Adopts a
        value of 1<br/>for normalised cumulative probability
        distributions</html>');
    set(h.amp5, 'Visible', 'On', 'String', 'y-offset', 'TooltipString', ...
        '<html><b>y-offset</b><br/>Value of y when t => Inf.  Typically 0.
        </html>','Position', [1.0 3.4615 11.2 1.0769]);
    set(h.cen1, 'Visible', 'On', 'String', 'mean tau 1', 'TooltipString', ...
        '<html><b>Exponential time constant</b><br/>The average time (tau)
        spent in a certain FRET<br/>state in [s].  1/tau is the decay rate in
        [s].</html>');
    set(h.cen2, 'Visible', 'On', 'String', 'mean tau 2', 'TooltipString', ...
        '<html><b>Exponential time constant</b><br/>The average time (tau)
        spent in a certain FRET<br/>state in [s].  1/tau is the decay rate in
        [s].</html>');
    set(h.cen3, 'Visible', 'On', 'String', 'mean tau 3', 'TooltipString', ...
        '<html><b>Exponential time constant</b><br/>The average time (tau)
        spent in a certain FRET<br/>state in [s].  1/tau is the decay rate in
        [s].</html>');
    set(h.cen4, 'Visible', 'On', 'String', 'mean tau 4', 'TooltipString', ...
        '<html><b>Exponential time constant</b><br/>The average time (tau)
        spent in a certain FRET<br/>state in [s].  1/tau is the decay rate in
        [s].</html>');
    set(h.cen5, 'Visible', 'Off');
    set(h.wid1, 'Visible', 'Off');
    set(h.wid2, 'Visible', 'Off');
    set(h.wid3, 'Visible', 'Off');
    set(h.wid4, 'Visible', 'On', 'String', 'x-offset', 'TooltipString', ...
        '<html><b>Exponential time constant</b><br/>The average time (tau)
        spent in a certain FRET<br/>state in [s].  1/tau is the decay rate in
        [s].</html>', ...
        'Position', [1.0 6.9230 11.2 1.0769]);
    set(h.wid5, 'Visible', 'Off');
    % show offset y/n?
    set(h.y_offset, 'Visible', 'On');
    set(h.x_offset, 'Visible', 'On');
    offset = get(h.y_offset, 'Value');
    if offset == 0
        set(h.edit_lower_y_offset, 'Visible', 'Off');
```

```
          set(h.edit_start_y_offset, 'Visible', 'Off');
          set(h.edit_upper_y_offset, 'Visible', 'Off');
     else
          set(h.edit_lower_y_offset, 'Visible', 'On');
          set(h.edit_start_y_offset, 'Visible', 'On');
          set(h.edit_upper_y_offset, 'Visible', 'On');
     end
     offset = get(h.x_offset, 'Value');
     if offset == 0
          set(h.edit_lower_x_offset, 'Visible', 'Off');
          set(h.edit_start_x_offset, 'Visible', 'Off');
          set(h.edit_upper_x_offset, 'Visible', 'Off');
     else
          set(h.edit_lower_x_offset, 'Visible', 'On');
          set(h.edit_start_x_offset, 'Visible', 'On');
          set(h.edit_upper_x_offset, 'Visible', 'On');
     end
     % adjust visibility of text fields
     set(h.lower_amp1, 'Visible', 'On', 'String', '0.01', ...
          'Enable', 'On');
     set(h.lower_amp2, 'Visible', 'Off', 'String', '0.01', ...
          'Enable', 'On');
     set(h.lower_amp3, 'Visible', 'Off', 'String', '0.01', ...
          'Enable', 'On');
     set(h.lower_amp4, 'Visible', 'Off', 'String', '0.01', ...
          'Enable', 'On');
     set(h.lower_amp5, 'Visible', 'Off', 'Enable', 'On');
     set(h.start_amp1, 'Visible', 'On', 'String', '0.8', ...
          'Enable', 'On');
     set(h.start_amp2, 'Visible', 'Off', 'String', '0.8', ...
          'Enable', 'On');
     set(h.start_amp3, 'Visible', 'Off', 'String', '0.8', ...
          'Enable', 'On');
     set(h.start_amp4, 'Visible', 'Off', 'String', '0.8', ...
          'Enable', 'On');
     set(h.start_amp5, 'Visible', 'Off', 'Enable', 'On');
     set(h.upper_amp1, 'Visible', 'On', 'String', '1', ...
          'Enable', 'On');
     set(h.upper_amp2, 'Visible', 'Off', 'String', '1', ...
          'Enable', 'On');
     set(h.upper_amp3, 'Visible', 'Off', 'String', '1', ...
          'Enable', 'On');
     set(h.upper_amp4, 'Visible', 'Off', 'String', '1', ...
          'Enable', 'On');
     set(h.upper_amp5, 'Visible', 'Off', 'Enable', 'On');
     set(h.lower_cen1, 'Visible', 'On', 'String', '0', ...
          'Enable', 'On');
     set(h.lower_cen2, 'Visible', 'Off', 'String', '0', ...
          'Enable', 'On');
     set(h.lower_cen3, 'Visible', 'Off', 'String', '0', ...
```

```
     ’Enable’, ’On’);
set(h.lower_cen4, ’Visible’, ’Off’, ’String’, ’0’, ...
     ’Enable’, ’On’);
set(h.lower_cen5, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.start_cen1, ’Visible’, ’On’, ’String’, ’50’, ...
     ’Enable’, ’On’);
set(h.start_cen2, ’Visible’, ’Off’, ’String’, ’50’, ...
     ’Enable’, ’On’);
set(h.start_cen3, ’Visible’, ’Off’, ’String’, ’50’, ...
     ’Enable’, ’On’);
set(h.start_cen4, ’Visible’, ’Off’, ’String’, ’50’, ...
     ’Enable’, ’On’);
set(h.start_cen5, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.upper_cen1, ’Visible’, ’On’, ’String’, ’Inf’, ...
     ’Enable’, ’On’);
set(h.upper_cen2, ’Visible’, ’Off’, ’String’, ’Inf’, ...
     ’Enable’, ’On’);
set(h.upper_cen3, ’Visible’, ’Off’, ’String’, ’Inf’, ...
     ’Enable’, ’On’);
set(h.upper_cen4, ’Visible’, ’Off’, ’String’, ’Inf’, ...
     ’Enable’, ’On’);
set(h.upper_cen5, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.lower_wid1, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.lower_wid2, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.lower_wid3, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.lower_wid4, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.lower_wid5, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.start_wid1, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.start_wid2, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.start_wid3, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.start_wid4, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.start_wid5, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.upper_wid1, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.upper_wid2, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.upper_wid3, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.upper_wid4, ’Visible’, ’Off’, ’Enable’, ’On’);
set(h.upper_wid5, ’Visible’, ’Off’, ’Enable’, ’On’);
% adjust dropdown menu
set(h.num_gauss, ’String’, {’single exp.  decay’;’double exp.  decay’;’triple
exp.  decay’;’quadruple exp.  decay’; ’stretched exp.  decay’}, ’Value’, 1);
% hide radiobuttons Gaussian fitting/thresholding
set(h.radiobutton_gaussfit, ’Visible’, ’Off’);
set(h.radiobutton_thresholding, ’Visible’, ’Off’);
% show equations
set(h.display_eq, ’Visible’, ’On’, ’String’, ’y = a*exp{-((x-b)/tau)} + c’);
% hide RMSE analysis
set(h.text29, ’Enable’, ’Off’);
set(h.text39, ’Enable’, ’Off’);
set(h.text33, ’Enable’, ’Off’);
set(h.text32, ’Enable’, ’Off’);
```

```
     set(h.text31, 'Enable', 'Off');
     set(h.text30, 'Enable', 'Off');
     set(h.edit_rms_penalty, 'Enable', 'Off');
     set(h.edit_rms_suggestion, 'Enable', 'Off');
     set(h.pushbutton_rms, 'Enable', 'Off');
     set(h.edit_rms_2gaussians, 'Enable', 'Off');
     set(h.edit_rms_3gaussians, 'Enable', 'Off');
     set(h.edit_rms_4gaussians, 'Enable', 'Off');
     set(h.edit_rms_5gaussians, 'Enable', 'Off');
end


% exit programme
function submenu_close_Callback(obj, evd, h)
close(h.bobafret);


% checkbox "x-offset"
% make edit field appear/disappear
function x_offset_Callback(obj, evd, h)
offset = get(h.x_offset, 'Value');
if offset == 0
     set(h.edit_lower_x_offset, 'Visible', 'Off');
     set(h.edit_start_x_offset, 'Visible', 'Off');
     set(h.edit_upper_x_offset, 'Visible', 'Off');
else
     set(h.edit_lower_x_offset, 'Visible', 'On');
     set(h.edit_start_x_offset, 'Visible', 'On');
     set(h.edit_upper_x_offset, 'Visible', 'On');
end


% checkbox "y-offset"
% make edit field appear/disappear
function y_offset_Callback(obj, evd, h)
offset = get(h.y_offset, 'Value');
if offset == 0
     set(h.edit_lower_y_offset, 'Visible', 'Off');
     set(h.edit_start_y_offset, 'Visible', 'Off');
     set(h.edit_upper_y_offset, 'Visible', 'Off');
else
     set(h.edit_lower_y_offset, 'Visible', 'On');
     set(h.edit_start_y_offset, 'Visible', 'On');
     set(h.edit_upper_y_offset, 'Visible', 'On');
end


% fitting parametres
% amplitudes
% amplitude 1, lower guess
```

```
function lower_amp1_Callback(obj, evd, h)
function lower_amp1_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 1, start guess
function start_amp1_Callback(obj, evd, h)
function start_amp1_CreateFcn(obj, evd, h)

if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 1, upper guess
function upper_amp1_Callback(obj, evd, h)
function upper_amp1_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 2, lower guess
function lower_amp2_Callback(obj, evd, h)
function lower_amp2_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 2, start guess
function start_amp2_Callback(obj, evd, h)
function start_amp2_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 2, upper guess
function upper_amp2_Callback(obj, evd, h)
function upper_amp2_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 3, lower guess
function lower_amp3_Callback(obj, evd, h)
```

```
function lower_amp3_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 3, start guess
function start_amp3_Callback(obj, evd, h)
function start_amp3_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 3, upper guess
function upper_amp3_Callback(obj, evd, h)
function upper_amp3_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 4, lower guess
function lower_amp4_Callback(obj, evd, h)
function lower_amp4_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 4, start guess
function start_amp4_Callback(obj, evd, h)
function start_amp4_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 4, upper guess
function upper_amp4_Callback(obj, evd, h)
function upper_amp4_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 5, start guess
function start_amp5_Callback(obj, evd, h)
function start_amp5_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
```

```
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 5, upper guess
function upper_amp5_Callback(obj, evd, h)
function upper_amp5_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% amplitude 5, lower guess
function lower_amp5_Callback(obj, evd, h)
function lower_amp5_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end


% centres
% centre 1, lower guess
function lower_cen1_Callback(obj, evd, h)
function lower_cen1_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% centre 1, start guess
function start_cen1_Callback(obj, evd, h)
function start_cen1_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% centre 1, upper guess
function upper_cen1_Callback(obj, evd, h)
function upper_cen1_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% centre 2, lower guess
function lower_cen2_Callback(obj, evd, h)
function lower_cen2_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
```

```
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% centre 2, start guess
function start_cen2_Callback(obj, evd, h)
function start_cen2_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% centre 2, upper guess
function upper_cen2_Callback(obj, evd, h)
function upper_cen2_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% centre 3, lower guess
function lower_cen3_Callback(obj, evd, h)
function lower_cen3_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% centre 3, start guess
function start_cen3_Callback(obj, evd, h)
function start_cen3_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% centre 3, upper guess
function upper_cen3_Callback(obj, evd, h)
function upper_cen3_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% centre 4, lower guess
function lower_cen4_Callback(obj, evd, h)
function lower_cen4_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
```

```
end

% centre 4, start guess
function start_cen4_Callback(obj, evd, h)
function start_cen4_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% centre 4, upper guess
function upper_cen4_Callback(obj, evd, h)
function upper_cen4_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% center 5, lower guess
function lower_cen5_Callback(obj, evd, h)
function lower_cen5_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% center 5, start guess
function start_cen5_Callback(obj, evd, h)
function start_cen5_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% center 5, upper guess
function upper_cen5_Callback(obj, evd, h)
function upper_cen5_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end


% width
% width 1, lower guess
function lower_wid1_Callback(obj, evd, h)
function lower_wid1_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
```

```
end

% width 1, start guess
function start_wid1_Callback(obj, evd, h)
function start_wid1_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 1, upper guess
function upper_wid1_Callback(obj, evd, h)
function upper_wid1_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 2, lower guess
function lower_wid2_Callback(obj, evd, h)
function lower_wid2_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 2, start guess
function start_wid2_Callback(obj, evd, h)
function start_wid2_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 2, upper guess
function upper_wid2_Callback(obj, evd, h)
function upper_wid2_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 3, lower guess
function lower_wid3_Callback(obj, evd, h)
function lower_wid3_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end
```

```
% width 3, start guess
function start_wid3_Callback(obj, evd, h)
function start_wid3_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 3, upper guess
function upper_wid3_Callback(obj, evd, h)
function upper_wid3_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 4, lower guess
function lower_wid4_Callback(obj, evd, h)
function lower_wid4_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 4, start guess
function start_wid4_Callback(obj, evd, h)
function start_wid4_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 4, upper guess
function upper_wid4_Callback(obj, evd, h)
function upper_wid4_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 5, upper guess
function upper_wid5_Callback(obj, evd, h)
function upper_wid5_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 5, start guess
function start_wid5_Callback(obj, evd, h)
```

```
function start_wid5_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% width 5, lower guess
function lower_wid5_Callback(obj, evd, h)
function lower_wid5_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end



% offset
% y-offset, lower guess
function edit_lower_y_offset_Callback(obj, evd, h)
function edit_lower_y_offset_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% y-offset, start guess
function edit_start_y_offset_Callback(obj, evd, h)
function edit_start_y_offset_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% y-offset, upper guess
function edit_upper_y_offset_Callback(obj, evd, h)
function edit_upper_y_offset_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% x-offset, upper guess
function edit_upper_x_offset_Callback(obj, evd, h)
function edit_upper_x_offset_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% y-offset, start guess
function edit_start_x_offset_Callback(obj, evd, h)
```

```
function edit_start_x_offset_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% y-offset, lower guess
function edit_lower_x_offset_Callback(obj, evd, h)
function edit_lower_x_offset_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), ...
          get(0,'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

function normalizeFig(h_fig)

children = get(h_fig, 'Children');
for i = 1:size(children,1)
     if isprop(children(i,1), 'Units')
          set(children(i,1), 'Units', 'normalized');
     end
     if isprop(children(i,1), 'Children')
          normalizeFig(children(i,1));
     end
end
```

### 12.3.2 clustTrans.m /written by Mélodie Hadzic)

#### Initialisation code

```
function varargout = clustTrans(varargin)
% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name', mfilename, ...
               'gui_Singleton', gui_Singleton, ...
               'gui_OpeningFcn', @clustTrans_OpeningFcn, ...
               'gui_OutputFcn', @clustTrans_OutputFcn, ...
               'gui_LayoutFcn', [] , ...
               'gui_Callback', []);
if nargin && ischar(varargin{1})
     gui_State.gui_Callback = str2func(varargin{1});
end
if nargout
     [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
     gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT
```

### Openingfunction

```
% -- Executes just before clustTrans is made visible.
function clustTrans_OpeningFcn(obj, evd,h, varargin)
clc;
setProp(get(obj, 'Children'), 'Units', 'normalized');
h.output = obj;
h.histarray = varargin{1};
guidata(obj,h);
setDefaultSetting(h.figure_clustTrans);
loadDat(h.figure_clustTrans);

% -- Outputs from this function are returned to the command line.
function varargout = clustTrans_OutputFcn(obj, evd,h)
varargout{1} = h.output;
```

### Callback

```
% ----------------Callback-------------------
function menu_expGraph_Callback(obj,evd,h)
axesDat = get(h.axes_2Dplot, 'UserData');
figure('Name', 'Summed 2D plots', 'Color', [1 1 1], 'NumberTitle', 'off', ...
'Units', 'normalized', 'OuterPosition', [0 0 1 1]);
colormap(colormap(h.axes_2Dplot));
h_axes = axes();
axesDat = get(h.axes_2Dplot, 'UserData');
isClust = size(axesDat,2) > 1;
raw_plot = axesDat{1};
Tdplot = raw_plot;
for j = 1:size(Tdplot,3)
     maxPlot = max(max(Tdplot(:,:,j)));
     if get(h.checkbox_gconv, 'Value')
          Tdplot(:,:,j) = convGauss(Tdplot(:,:,j));
          maxPlot_gauss = max(max(Tdplot(:,:,j)));
     end
     if get(h.checkbox_gconv, 'Value')
          maxPlot_n = maxPlot_gauss;
     else
          maxPlot_n = maxPlot;
     end
     if get(h.checkbox_norm, 'Value')
          if ~maxPlot_n
               maxPlot_n = 1;
          end
          Tdplot(:,:,j) = Tdplot(:,:,j)/maxPlot_n;
     end
end
if size(Tdplot,3) == 3
     [r,c,o] = find(~Tdplot(:,:,1) & ~Tdplot(:,:,2) & ~Tdplot(:,:,3));
     for i = 1:size(r,1)
```

```
          Tdplot(r(i,1),c(i,1),:)  = 1;
          end
end
imagesc(h.param.xLim, h.param.yLim, Tdplot, 'Parent', h_axes);
axis(h_axes, 'image');
xlabel(h_axes, 'Initial FRET');
ylabel(h_axes, 'Next FRET');
if get(h.checkbox_norm,'Value')
     ylabel(h.colorbar, 'normalized occurence');
else
     ylabel(h.colorbar, 'occurence');
end
grid(h_axes, 'on');
set(h_axes, 'XAxisLocation', 'top', 'YAxisLocation', 'right', ...
          'XLim', h.param.xLim, 'YLim', h.param.yLim);
h.colorbar = colorbar('peer', h_axes);
if get(h.checkbox_norm,'Value')
     ylabel(h.colorbar, 'normalized occurence');
else
     ylabel(h.colorbar, 'occurence');
end
set(h_axes,'NextPlot','add');

if isClust
     colClust = h.param.start.colClust;
     pnt_clust = axesDat{2};
     leg = {};
     for i = 1:h.param.nbTrans
          plot(h_axes, pnt_clust{i}(:,1), pnt_clust{i}(:,2), 'LineStyle', ...
          'none', 'Marker', '.', 'Color', h.param.colList(colClust(i),:));
          leg = {leg{:}, ['cluster n:°' num2str(i)]};
     end
     legend(h_axes, leg);
end

if isfield(h.param, 'start') && isfield(h.param.start, 'iniFRET')
     for i = 1:h.param.nbTrans
          plot(h_axes, h.param.start.iniFRET(i), h.param.start.finFRET(i), ...
               'LineStyle', 'none', 'LineWidth', 3, 'Marker', '+', 'MarkerSize', ...
               15, 'Color', h.param.colList(h.param.start.colClust(i), :));
     end
end
plotGrid(h_axes);
set(h_axes,'NextPlot','replacechildren');
```

### Main functions

```
% -- Executes on button press in pushbutton_update.
function pushbutton_update_Callback(obj, evd, h)
```

```
if h.param.nbTrans
     ok = checkStartParam(h.param, h.figure_clustTrans);
     if ok
     pushbutton_reset_Callback(h.pushbutton_reset, evd, h);
          [means_cvg pnt_clust occ] = cluster_tr(h.figure_clustTrans);
          h.param.result.iniFRET(1:h.param.nbTrans) = means_cvg(:,1)’;
          h.param.result.finFRET(1:h.param.nbTrans) = means_cvg(:,2)’;
          h.param.result.cluster = pnt_clust;
          h.param.result.occ = occ;
          guidata(h.figure_clustTrans, h);
          plot2Dplot(h.figure_clustTrans, get(h.edit_contPan, ’String’));
          curTrans = get(h.listbox_trans,’Value’);
          set(h.edit_iniFRET_res, ’String’, num2str(h.param.result.iniFRET(
               curTrans)));
          set(h.edit_finFRET_res, ’String’, num2str(h.param.result.finFRET(
          curTrans)));
          set(h.text_nbClustTrans, ’String’, [num2str(100*h.param.result.occ(
          curTrans))]);
          set(h.pushbutton_ok, ’Enable’, ’on’);
     end
else
     setContPan(’Error:  The number of transition must be a positive integer !’, ...
          ’error’, h.figure_clustTrans);
end

% -- Executes on button press in pushbutton_reset.
function pushbutton_reset_Callback(obj, evd, h)
if isfield(h.param, ’result’)
     h.param = rmfield(h.param, ’result’);
     guidata(h.figure_clustTrans, h);
end
set(h.edit_iniFRET_res, ’String’, ”);
set(h.edit_finFRET_res, ’String’, ”);
set(h.text_nbClustTrans, ’String’, ”);
set(h.pushbutton_ok, ’Enable’, ’off’);
plot2Dplot(h.figure_clustTrans, ’Transitions 2D plot has been successfully
     refreshed !’);
% -- Executes on button press in checkbox_gconv.
function checkbox_gconv_Callback(obj, evd, h)
plot2Dplot(h.figure_clustTrans, ’Transitions 2D plot has been successfully
     refreshed !’);
% -- Executes on button press in checkbox_norm.
function checkbox_norm_Callback(obj, evd, h)
plot2Dplot(h.figure_clustTrans, ’Transition density plot has been successfully
     normalized !’);
% -- Executes on selection change in listbox_trans.
function listbox_trans_Callback(obj, evd,h)
curTrans = get(obj,’Value’);
if isfield(h.param, ’start’) && size(h.param.start.iniFRET,2) >= curTrans &&
     ~isempty(h.param.start.iniFRET(curTrans))
```

```
    set(h.edit_iniFRET_start, ’String’, num2str(h.param.start.iniFRET(curTrans)));
    set(h.edit_finFRET_start, ’String’, num2str(h.param.start.finFRET(curTrans)));
    set(h.edit_radius, ’String’, num2str(h.param.start.radius(curTrans)));
    set(h.popupmenu_color, ’Value’, h.param.start.colClust(curTrans));
    set(h.edit_color, ’BackgroundColor’, h.param.colList(h.param.start.colClust(
        curTrans),:));
else
    set(h.edit_iniFRET_start, ’String’, ’0’);
    set(h.edit_finFRET_start, ’String’, ’0’);
    set(h.edit_radius, ’String’, ’1’);
    set(h.popupmenu_color, ’Value’, ’1’);
    set(h.edit_color, ’BackgroundColor’, h.param.colList(1,:));
end
if isfield(h.param, ’result’) && size(h.param.result.iniFRET,2) >= curTrans &&
    ∼isempty(h.param.result.iniFRET(curTrans))
    set(h.edit_iniFRET_res, ’String’, num2str(h.param.result.iniFRET(curTrans)));
    set(h.edit_finFRET_res, ’String’, num2str(h.param.result.finFRET(curTrans)));
    set(h.text_nbClustTrans, ’String’, num2str(size(h.param.result.cluster{
        curTrans},1)));
    set(h.text_nbClustTrans, ’String’, [num2str(100*h.param.result.occ(curTrans))]);
else
    set(h.edit_iniFRET_res, ’String’, ”);
    set(h.edit_finFRET_res, ’String’, ”);
    set(h.text_nbClustTrans, ’String’, ”);
end
drawnow;
plot2Dplot(h.figure_clustTrans, ’Transitions 2D plot has been successfully
    refreshed !’);
function edit_nbTrans_Callback(obj, evd,h)
pushbutton_reset_Callback(h.pushbutton_reset, evd, h);
h.param.nbTrans = round(str2num(get(obj,’String’)));
set(obj,’String’, num2str(h.param.nbTrans));
str = {};
for i = 1:h.param.nbTrans
    str = {str{:} [’transition n°’ num2str(i)]};
    if ∼isfield(h.param, ’start’) || size(h.param.start.iniFRET,2) < i
        h.param.start.iniFRET(i) = 0;
        h.param.start.finFRET(i) = 0;
        h.param.start.radius(i) = 1;
        if i <= size(h.param.colList,1)
            h.param.start.colClust(i) = i;
        else
            h.param.start.colClust(i) = size(h.param.colList,1);
        end
    end
end
guidata(h.figure_clustTrans,h);
set(h.listbox_trans, ’String’, str, ’Value’, 1);
listbox_trans_Callback(h.listbox_trans, evd,h)
```

```
function edit_xBin_Callback(obj, evd, h)
val = round(str2num(get(obj, 'String')));
if val < 1
     setContPan('Error:  x binning must be higher than 1', 'error', ...
          h.figure_clustTrans);
     val = h.param.xbin;
else
     h.param.xbin = val;
     guidata(h.figure_clustTrans, h);
end
set(obj, 'String', num2str(val));

function edit_yBin_Callback(obj, evd, h)
val = round(str2num(get(obj, 'String')));
if val < 1
     setContPan('Error:  y binning must be higher than 1', 'error', ...
          h.figure_clustTrans);
     val = h.param.ybin;
else
     h.param.ybin = val;
     guidata(h.figure_clustTrans, h);
end
set(obj, 'String', num2str(val));

function edit_xLim_low_Callback(obj, evd, h)
h.param.xLim = [str2num(get(obj, 'String')) h.param.xLim(2)];
guidata(h.figure_clustTrans, h);

function edit_xLim_up_Callback(obj, evd, h)
h.param.xLim = [h.param.xLim(1) str2num(get(obj, 'String'))];
guidata(h.figure_clustTrans, h);

function edit_yLim_low_Callback(obj, evd, h)
h.param.yLim = [str2num(get(obj, 'String')) h.param.yLim(2)];
guidata(h.figure_clustTrans, h);

function edit_yLim_up_Callback(obj, evd, h)
h.param.yLim = [h.param.yLim(1) str2num(get(obj, 'String'))];
guidata(h.figure_clustTrans, h);

function edit_iniFRET_start_Callback(obj, evd,h)
curTrans = get(h.listbox_trans,'Value');
h.param.start.iniFRET(curTrans) = str2num(get(obj,'String'));
set(obj, 'String', num2str(h.param.start.iniFRET(curTrans)));
guidata(h.figure_clustTrans,h);

function edit_finFRET_start_Callback(obj, evd,h)
curTrans = get(h.listbox_trans,'Value');
h.param.start.finFRET(curTrans) = str2num(get(obj,'String'));
set(obj, 'String', num2str(h.param.start.finFRET(curTrans)));
```

```
guidata(h.figure_clustTrans,h);

function edit_radius_Callback(obj, evd,h)
curTrans = get(h.listbox_trans,'Value');
h.param.start.radius(curTrans) = str2num(get(obj,'String'));
set(obj, 'String', num2str(h.param.start.radius(curTrans)));
guidata(h.figure_clustTrans,h);

% -- Executes on selection change in popupmenu_color.
function popupmenu_color_Callback(obj, evd,h)
set(h.edit_color, 'BackgroundColor', h.param.colList(get(obj,'Value'),:));
curTrans = get(h.listbox_trans,'Value');
h.param.start.colClust(curTrans) = get(obj,'Value');
guidata(h.figure_clustTrans,h);

% -- Executes on button press in pushbutton_colormap.
function pushbutton_colormap_Callback(obj, evd,h)
colormapeditor;
cmap = colormap;
setContPan('Colormap of the 2D plot has been successfully updated !', ...
     'success', h.figure_clustTrans);

function edit_maxIter_Callback(obj, evd, h)
h.param.maxIter = round(str2num(get(obj,'String')));
set(obj, 'String', num2str(h.param.maxIter));
guidata(h.figure_clustTrans,h);

% -- Executes on button press in pushbutton_updatePlot.
function pushbutton_updatePlot_Callback(obj, evd, h)
buildTDP(h.figure_clustTrans);
plot2Dplot(h.figure_clustTrans, 'Transition density plot has been successfully ...
     refreshed !');

% ---------------ButtonDownFcn-----------------
function axes_2Dplot_ButtonDownFcn(obj, evd, h_fig)
h = guidata(h_fig);
if h.param.nbTrans > 0
     currTrans = get(h.listbox_trans, 'Value');
     newPnt = get(h.axes_2Dplot, 'CurrentPoint');
     bin_x = (h.param.xLim(2) - h.param.xLim(1)) / h.param.xbin;
     intv_x = h.param.xLim(1):bin_x:h.param.xLim(2);
     bin_y = (h.param.yLim(2) - h.param.yLim(1)) / h.param.ybin;
     intv_y = h.param.yLim(1):bin_y:h.param.yLim(2);
     [o,indX,o] = find(intv_x <= newPnt(1,1));
     newPnt(1,1) = mean([intv_x(indX(numel(indX))) intv_x(indX(numel(indX))+1)]);
     [o,indY,o] = find(intv_y <= newPnt(1,2));
     newPnt(1,2) = mean([intv_y(indY(numel(indY))) intv_y(indY(numel(indY))+1)]);
     h.param.start.iniFRET(currTrans) = newPnt(1,1);
     h.param.start.finFRET(currTrans) = newPnt(1,2);
     set(h.edit_iniFRET_start, 'String', num2str(newPnt(1,1)));
```

```
     set(h.edit_finFRET_start, 'String', num2str(newPnt(1,2)));
     guidata(h_fig, h);
     plot2Dplot(h.figure_clustTrans, ['Transition n°:'  num2str(currTrans)
          ' updated to ' num2str(newPnt(1,1)) ' -> ' num2str(newPnt(1,2)) ' !']);
end

% ---------------CreateFcn-----------------
% -- Executes during object creation, after setting all properties.
function edit_iniFRET_res_CreateFcn(obj, evd,h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_finFRET_res_CreateFcn(obj, evd,h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function listbox_trans_CreateFcn(obj, evd,h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_nbTrans_CreateFcn(obj, evd,h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_iniFRET_start_CreateFcn(obj, evd,h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_finFRET_start_CreateFcn(obj, evd,h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
```

```
function edit_radius_CreateFcn(obj, evd,h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function popupmenu_color_CreateFcn(obj, evd,h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_xBin_CreateFcn(obj, evd,h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_yBin_CreateFcn(obj, evd,h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_contPan_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
     'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_color_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_maxIter_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
          'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_xLim_low_CreateFcn(obj, evd, h)
```

```
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
         'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_xLim_up_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
         'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_yLim_low_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
         'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% -- Executes during object creation, after setting all properties.
function edit_yLim_up_CreateFcn(obj, evd, h)
if ispc && isequal(get(obj,'BackgroundColor'), get(0, ...
         'defaultUicontrolBackgroundColor'))
     set(obj,'BackgroundColor','white');
end

% --------------Useful---------------

function loadDat(h_fig)
% Load 2D plot data from file, build the summed 2D plot and plot it
h = guidata(h_fig);
dat = h.histarray;
[discrVal,o,r] = unique(dat(:,2:3), 'rows');
h.param.xLim = [min(min(discrVal(:,1))) max(max(discrVal(:,1)))];
% previous value
h.param.yLim = [min(min(discrVal(:,2))) max(max(discrVal(:,2)))];
% next value
set(h.edit_xLim_low, 'String', num2str(h.param.xLim(1)));
set(h.edit_xLim_up, 'String', num2str(h.param.xLim(2)));
set(h.edit_yLim_low, 'String', num2str(h.param.yLim(1)));
set(h.edit_yLim_up, 'String', num2str(h.param.yLim(2)));
guidata(h_fig, h);
buildTDP(h_fig);
childFig = get(h_fig, 'Children');
setProp(childFig, 'Enable', 'on');
set(h.pushbutton_ok, 'Enable', 'off');
plot2Dplot(h_fig, 'Transition density plot has been successfully plotted !');

function buildTDP(h_fig)
```

```
% bin state values
str = 'Process:  (1/2) binning the discrete values ...';
setContPan(str, 'process', h_fig);
h = guidata(h_fig);
bin_x = (h.param.xLim(2) - h.param.xLim(1)) / h.param.xbin;
iv_x = h.param.xLim(1):bin_x:h.param.xLim(2);
bin_y = (h.param.yLim(2) - h.param.yLim(1)) / h.param.ybin;
iv_y = h.param.yLim(1):bin_y:h.param.yLim(2);
dat = h.histarray;
for i = 1:h.param.xbin-2
     dat((dat(:,2) >= iv_x(i) & dat(:,2) < iv_x(i+1)),2) = mean(iv_x(i:i+1));
     setContPan({str; [num2str(round(100*i/(h.param.xbin-2+h.param.ybin-2)))
     % 'completed']}, 'process', h_fig);
end
dat((dat(:,2) >= iv_x(h.param.xbin-1)),2) = mean([iv_x(h.param.xbin-1), ...
     iv_x(h.param.xbin)]);
for i = 1:h.param.ybin-2
     dat((dat(:,3) >= iv_y(i) & dat(:,3) < iv_y(i+1)),3) = mean(iv_y(i:i+1));
     setContPan({str; [num2str(round(100*(i+h.param.xbin-2)/ ...
          (h.param.xbin-2+h.param.ybin-2)))
     % 'completed']}, 'process', h_fig);
end
dat((dat(:,3) >= iv_y(h.param.ybin-1)),3) = mean([iv_y(h.param.ybin-1), ...
     iv_y(h.param.ybin)]);
[discrVal,o,r] = unique(dat(:,2:3), 'rows');
pnt_clust = cell(1,size(discrVal,1));
for j = 1:size(discrVal,1)
     [r2,o,o] = find(r == j);
     pnt_clust{j} = dat(r2',2:3);
end
% build the 2D plot FRETi-FRETf
str = 'Process:  (2/2) building the transition 2D plot...';
setContPan(str, 'process', h_fig);
TD_Dat_tot = zeros(h.param.ybin, h.param.xbin);
for j = 1:size(pnt_clust,2)
     for k = 1:size(pnt_clust{j},1)
     [o, valSupIni] = find(iv_x <= pnt_clust{j}(k,1));
     [o, valSupFin] = find(iv_y <= pnt_clust{j}(k,2));
     if ~(pnt_clust{j}(k,1) == pnt_clust{j}(k,2))
               TD_Dat_tot(valSupFin(size(valSupFin,2)),valSupIni(size(valSupIni,...
                    2))) = TD_Dat_tot(valSupFin(size(valSupFin,2)), ...
                    valSupIni(size(valSupIni,2))) + 1;
          end
     end
     setContPan({str; [num2str(round(100*j/size(pnt_clust,2)))
     % 'completed']}, 'process', h_fig);
end
axesData{1} = TD_Dat_tot;
set(h.axes_2Dplot, 'UserData', axesData);
function plot2Dplot(h_fig, str)
```

```
h = guidata(h_fig);
axesDat = get(h.axes_2Dplot, ’UserData’);
Tdplot = axesDat{1};
for j = 1:size(Tdplot,3)
     maxPlot = max(max(Tdplot(:,:,j)));
     maxPlot_n = maxPlot;
     if get(h.checkbox_gconv, ’Value’)
          Tdplot(:,:,j) = convGauss(Tdplot(:,:,j));
          maxPlot_gauss = max(max(Tdplot(:,:,j)));
          maxPlot_n = maxPlot_gauss;
     end
     if get(h.checkbox_norm, ’Value’)
          if ∼maxPlot_n
               maxPlot_n = 1;
          end
          Tdplot(:,:,j) = Tdplot(:,:,j)/maxPlot_n;
     end
end
if size(Tdplot,3) == 3
     [r,c,o] = find(∼Tdplot(:,:,1) & ∼Tdplot(:,:,2) & ∼Tdplot(:,:,3));
     for i = 1:size(r,1)
          Tdplot(r(i,1),c(i,1),:)  = 1;
     end
end
% plot
im = imagesc(h.param.xLim, h.param.yLim, Tdplot, ’Parent’, h.axes_2Dplot);
axis(h.axes_2Dplot, ’image’);
xlabel(h.axes_2Dplot, ’Initial FRET’);
ylabel(h.axes_2Dplot, ’Next FRET’);
set(im, ’ButtonDownFcn’, {@axes_2Dplot_ButtonDownFcn, h_fig});
h.colorbar = colorbar(’peer’, h.axes_2Dplot);
if get(h.checkbox_norm,’Value’)
     ylabel(h.colorbar, ’normalized occurence’);
else
     ylabel(h.colorbar, ’occurence’);
end
grid(h.axes_2Dplot, ’on’);
set(h.axes_2Dplot, ’NextPlot’, ’add’);
if isfield(h.param, ’result’) && isfield(h.param.result, ’iniFRET’) &&
          ∼isempty(h.param.result.iniFRET)
     leg = {};
     for i = 1:h.param.nbTrans
          plot(h.param.result.cluster{i}(:,1), h.param.result.cluster{i}(:,2), ...
               ’LineStyle’, ’none’, ’Marker’, ’.’, ’Color’, ...
               h.param.colList(h.param.start.colClust(i),:));
          leg = {leg{:}, [’cluster n:°’ num2str(i)]};
     end
     legend(h.axes_2Dplot, leg);
     axesDat = get(h.axes_2Dplot, ’UserData’);
     axesDat{2} = h.param.result.cluster;
```

```
     set(h.axes_2Dplot, ’UserData’, axesDat);
     for i = 1:h.param.nbTrans
          plot(h.axes_2Dplot, h.param.result.iniFRET(i), ...
               h.param.result.finFRET(i), ’LineStyle’, ’none’, ’LineWidth’, ...
               3, ’Marker’, ’+’, ’MarkerSize’, 15, ...
               ’Color’, h.param.colList(h.param.start.colClust(i), :));
     end
     set(h.axes_2Dplot, ’NextPlot’, ’replacechildren’);
elseif isfield(h.param, ’start’) && isfield(h.param.start, ’iniFRET’)
     for i = 1:h.param.nbTrans
          plot(h.axes_2Dplot, h.param.start.iniFRET(i), ...
               h.param.start.finFRET(i), ’LineStyle’, ’none’, ’LineWidth’, ...
               3, ’Marker’, ’+’, ’MarkerSize’, 15, ’Color’,
               h.param.colList(h.param.start.colClust(i), :));
     end
end
guidata(h_fig, h);
plotGrid(h.axes_2Dplot);
set(h.axes_2Dplot, ’XAxisLocation’, ’top’, ’YAxisLocation’, ’right’, ...
               ’UserData’, axesDat, ’ButtonDownFcn’, ...
               {@axes_2Dplot_ButtonDownFcn, h_fig}, ’XLim’, h.param.xLim, ...
               ’YLim’, h.param.yLim, ’YDir’’normal’, ’CLim’, ...
               [min(min(Tdplot))-(max(max(Tdplot))-min(min(Tdplot)))/ ...
               (size(colormap,1)-1) max(max(Tdplot))+(max(max(Tdplot)) - ...
               min(min(Tdplot)))/(size(colormap,1)-1)]);
set(h.axes_2Dplot, ’NextPlot’, ’replacechildren’);
setContPan(str, ’success’, h_fig);
setProp([h.edit_color;h.edit_iniFRET_res;h.edit_finFRET_res], ’Enable’, ’inactive’);

function setDefaultSetting(h_fig)
% Set default parameters at the ConcDt tool opening
h = guidata(h_fig);
% Default colormap -> 256 color
blue = [1 1/85:1/85:1 (1-1/85):-1/85:0 zeros(1,85)]’;
green = [1 zeros(1,85) 1/85:1/85:1 (1-1/85):-1/85:0]’;
red = [1 zeros(1,170) 1/85:1/85:1]’;
cmap = [red green blue];
setCmap(h_fig, cmap);
if ∼isfield(h, ’param’) || ∼isfield(h.param, ’xLim’) || isempty(h.param.xLim)
     h.param.xbin = 200;
     h.param.ybin = 200;
     h.param.xLim = [-0.2 1.2];
     h.param.yLim = [-0.2 1.2];
     h.param.colList = [1 0 0
                    0 1 0
                    0 0 1
                    1 1 0
                    0 1 1
                    1 0 1
                    0 0.5 0
```

```
                    1 0.5 0
                    0.5 0 0
                    0 0.5 0
                    0 0 0.5
                    0.5 0.5 0
                    0 0.5 0.5
                    0.5 0 0.5
                    0.5 0.25 0
                    1 0.5 1
                    0.5 0 1
                    0.5 0.5 0.5
                    1 1 0.5
                    1 0.5 1
                    0.5 1 1];
     h.param.nbTrans = 2;
     h.param.maxIter = 50;
     h.param.start.iniFRET(1) = 0;
     h.param.start.finFRET(1) = 0;
     h.param.start.iniFRET(2) = 0;
     h.param.start.finFRET(2) = 0;
     h.param.start.radius(1) = 1;
     h.param.start.radius(2) = 1;
     h.param.start.colClust(1) = 1;
     h.param.start.colClust(2) = 2;
end
guidata(h_fig, h);
set(h.edit_xBin, 'String', num2str(h.param.xbin));
set(h.edit_yBin, 'String', num2str(h.param.ybin));
set(h.edit_xLim_low, 'String', num2str(h.param.xLim(1)));
set(h.edit_xLim_up, 'String', num2str(h.param.xLim(2)));
set(h.edit_yLim_low, 'String', num2str(h.param.yLim(1)));
set(h.edit_yLim_up, 'String', num2str(h.param.yLim(2)));
set(h.edit_nbTrans, 'String', num2str(h.param.nbTrans));
set(h.edit_maxIter, 'String', num2str(h.param.maxIter));
str = {};
for i = 1:h.param.nbTrans
     str = {str{:}, ['transition n°:  num2str(i)]};
end
set(h.listbox_trans, 'String', str, 'Value', 1);
set(h.edit_iniFRET_start, 'String', num2str(h.param.start.iniFRET(1)));
set(h.edit_finFRET_start, 'String', num2str(h.param.start.finFRET(1)));
set(h.edit_radius, 'String', num2str(h.param.start.radius(1)));

function setCmap(h_fig, cmap)
h = guidata(h_fig);
colormap(cmap);
cmapFin = 0:1/99:1;
imagesc(cmapFin, 'Parent', h.axes_colormap);
set(h.axes_colormap, 'YTick', [], 'XLim', [0 100]);
```

```
function ok = checkStartParam(param, h_fig)
% Check if the values of all starting parameters are correct
ok = 0;
if (isfield(param, 'start') && ...
     sum(isfield(param.start, {'iniFRET', ...
                               'finFRET', ...
                               'radius', ...
                               'colClust'}), 2))
     if (size(param.start.iniFRET,2) >= param.nbTrans && ...
          size(param.start.finFRET,2) >= param.nbTrans && ...
          size(param.start.radius,2) >= param.nbTrans && ...
          size(param.start.colClust,2) >= param.nbTrans)
          ok = 1;
          startDat = [param.start.iniFRET' param.start.finFRET'];
          for i = 1:param.nbTrans
               [r1,o,o] = find(startDat(:,1) == startDat(i,1) & startDat(:,2)
                    == startDat(i,2));
               if ~isempty(r1)
                    for j = 1:size(r1,1)
                         if r1(j,1) ~= i
                              setContPan('Error:  the starting transition
                              coordinates must be differents.', 'error', ...
                              h_fig);
                              ok = 0;
                              break;
                         end
                    end
               end
          end
     else
          setContPan('Error:  all starting parameters for each transition have
               to be set !', 'error', h_fig);
     end
else
     setContPan('Error:  all starting parameters for each transition have to be
          set !', 'error', h_fig);
end

function [means_cvg pnt_clust occ] = cluster_tr(h_fig)
h = guidata(h_fig);
nbTrans = h.param.nbTrans;
iniFRET = h.param.start.iniFRET(1:nbTrans);
finFRET = h.param.start.finFRET(1:nbTrans);
radius = h.param.start.radius(1:nbTrans);
axesDat = get(h.axes_2Dplot, 'UserData');
TD_Dat_tot = axesDat{1};
[means_cvg pnt_clust occ] = getKmean([iniFRET' finFRET'], radius', ...
     TD_Dat_tot, h_fig);
occ = occ/sum(occ);
function dataConv = convGauss(data)
```

```
gFilter = exp(-((-5:5)/2).2);
gFilter = gFilter/sum(gFilter,2);
dataConv = conv2(data',gFilter','same');
dataConv = conv2(dataConv',gFilter','same');

function pushbutton_ok_Callback(obj, evd, h)
h = guidata(h.figure_clustTrans);
nbTrans = h.param.nbTrans;
pnt_clust = h.param.result.cluster;
trans = [h.param.result.iniFRET' h.param.result.finFRET'];
str = ['Rebuild the dwell-times list from the ' num2str(h.param.nbTrans)
    ' clusters ...'];
setContPan(str, 'process', h.figure_clustTrans);
dtDat = h.histarray;
h.histarray = [];
dtDat = binValues(dtDat, trans, pnt_clust, h.figure_clustTrans);
n = 0;
if size(dtDat,1) > 2
    for ii = 1:nbTrans
        [r,o,o] = find(dtDat(:,2) == trans(ii,1) & dtDat(:,3) == trans(ii,2));
        if ~isempty(r)
            h.histarray((size(h.histarray,1)+1):(size(h.histarray,1)+...
                size(r,1)),:)  = dtDat(r',:);
        else
            n = n + 1;
            str = [num2str(n) '/' num2str(size(dtDat,1)) ' transitions exclude
                from the list ...'];
            setContPan(str, 'error', h.figure_clustTrans);
        end
    end
else
    disp('No or only one transition detected:  ');
end
h.transitions = trans;
guidata(h.figure_clustTrans, h);
set(h.figure_clustTrans, 'Visible', 'off');
uiresume;

function dt_final = binValues(dt, trans, pnt_clust, h_fig)
h = guidata(h_fig);
str = 'Process:  (1/3) binning the discrete values ...';
setContPan(str, 'process', h_fig);
% assign original FRET states to bin values
bin_x = (h.param.xLim(2) - h.param.xLim(1)) / h.param.xbin;
iv_x = h.param.xLim(1):bin_x:h.param.xLim(2);
bin_y = (h.param.yLim(2) - h.param.yLim(1)) / h.param.ybin;
iv_y = h.param.yLim(1):bin_y:h.param.yLim(2);
for i = 1:h.param.xbin-2
    dt((dt(:,2) >= iv_x(i) & dt(:,2) < iv_x(i+1)),2) = mean(iv_x(i:i+1));
    setContPan({str; [num2str(round(100*i/(h.param.xbin-2+h.param.ybin-2))) '
```

```
     % completed']}, 'process', h_fig);
end
dt((dt(:,2) >= iv_x(h.param.xbin-1)),2) = mean([iv_x(h.param.xbin-1), ...
     iv_x(h.param.xbin)]);
for i = 1:h.param.ybin-2
     dt((dt(:,3) >= iv_y(i) & dt(:,3) < iv_y(i+1)),3) = mean(iv_y(i:i+1));
     setContPan({str; [num2str(round(100*(i+h.param.xbin-2)/(h.param.xbin- ...
          2+h.param.ybin-2)))
     % 'completed']}, 'process', h_fig);
end
dt((dt(:,3) >= iv_y(h.param.ybin-1)),3) = mean([iv_y(h.param.ybin-1), ...
     iv_y(h.param.ybin)]);
dt_stateBin = dt;
str = 'Process:  (2/3) assign dwell-times to respective clusters ...';
setContPan(str, 'process', h_fig);
n = 0;
for j = 1:size(trans,1)
     for i = 1:size(pnt_clust{j},1)
          isPnt = ismember(dt_stateBin(:,2:3), pnt_clust{j}(i,1:2), 'rows');
          dt_stateBin(isPnt',2) = trans(j,1);
          dt_stateBin(isPnt',3) = trans(j,2);
          n = n+1;
          setContPan({str; [num2str(round(100*n/(size(trans,1)* ...
               size(pnt_clust{j},1))))
     % 'completed']}, 'process', h_fig);
     end
end
dt_final = dt_stateBin;
% if the dwell-times array contain the file number (first and last
% transitions detected known)
if size(dt_stateBin,2) > 3
     dt_final = [];
     str = 'Process:  (3/3) sum up dwell-times of successive "static"
          transitions ...';
     setContPan(str, 'process', h_fig);
     mol_nb = unique(dt_stateBin(:,size(dt_stateBin,2)));
     for m = 1:size(mol_nb,1)
          dt_stateBin_mol = dt_stateBin((dt_stateBin(:, ...
               size(dt_stateBin,2)) == m),:);
          [r,o,o] = find((dt_stateBin_mol(:,2) == dt_stateBin_mol(:,3)));
          % find A->A transitions
          r_double = [];
          if ~isempty(r)
               [r_double,o,o] = find(r == [r(2:size(r,1),1) - 1;0]);
               % find the index in r of rows being the first of a multiple
               % A->A row series
               dt_final_mol = dt_stateBin_mol(1:r(1)-1,:);
               for i = 1:size(r,1)-1
                    % for each A->A row
                    if ~isempty(r_double) && ~isempty(find(r_double==i))
```

```
                    % being succeded by a same other A->A row
                         dt_stateBin_mol(r(i)+1, 1) = dt_stateBin_mol(r(i)+1, ...
                              1) + dt_stateBin_mol(r(i), 1);
                    else
                         dt_final_mol(size(dt_final_mol,1)+1,1:3) = [
                              dt_stateBin_mol(r(i), 1) + dt_stateBin_mol(
                              r(i)+1, 1) dt_stateBin_mol(r(i)+1, 2:3)];
                         if r(i+1)- r(i) > 2
                              dt_final_mol(size(dt_final_mol,1)+1:(size(
                                   dt_final_mol,1)+r(i+1)- r(i) - 2),1:3) = ...
                                   dt_stateBin_mol(r(i)+2:r(i+1)-1,:);
                         end
                    end
                    setContPan({str; [num2str(round(100*i/(size(r,1)-1))) '
                    % completed']}, 'process', h_fig);
               end
               if size(dt_stateBin_mol,1) >= r(numel(r),1)+1
                    dt_final_mol(size(dt_final_mol,1)+1,:)  = ...
                         [dt_stateBin_mol(r(numel(r),1),1)+ ...
                         dt_stateBin_mol(r(numel(r),1)+1,1)
                         dt_stateBin_mol(r(numel(r),1)+1,2:3)];
               end
               if size(dt_stateBin_mol,1) >= r(numel(r),1)+2
                    dt_final_mol = [dt_final_mol dt_stateBin_mol(r(numel(r),1)+ ...
                         2:size(dt_stateBin_mol,1),:)];
               end
          else
               dt_final_mol = dt_stateBin_mol;
          end
          if size(dt_final_mol,1) > 2 && size(dt_final_mol,2) > 3
               % delete the first and last dwell-times (exact duration of
               % previous state unknown)
               dt_final_mol = dt_final_mol(2:(size(dt_final_mol,1)-1),:);
          end
          dt_final = [dt_final dt_final_mol];
     end
end

function setProp(h, prop, val)
for i = 1:numel(h)
     if isprop(h(i),'Children')
          h_obj = get(h(i), 'Children');
          if ∼isempty(h_obj)
               setProp(h_obj, prop, val);
          end
     end
     if isprop(h(i),prop)
          set(h(i), prop, val);
     end
end
```

### 12.3.3 getKman.m (written by Mélodie Hadzic)

```
function [means_cvg clusters I_clust] = getKmean(means_start, tol, data, varargin)
% --------------------------------------------------
%| |
%| [means_cvg clusters I_clust] = getKmean(means_start, tol, data, varargin) |
%| |
% --------------------------------------------------
%| |
%| getKmean determines the centres of the k clusters from a set of data points |
%| using a weighted k-mean algorithm.  |
%| |
%| The algorithm needs an initial guess of the (x,y) coordinates of |
%| the cluster centres.  |
%| |
%| The weight given to each data point in a cluster is the ratio of its |
%| intensity (e.g occurence) on the distance from the current center of the |
%| cluster.  |
%| |
%| - INPUT ------------------------------------------- |
%| |
%| means_start:  nx2 matrix containing the intial (x,y) guess of the n centres |
%| |
%| tol:  nx1 vector containing the tolerance radius of the n clusters |
%| |
%| data:  axb matrix mapping the intensity of the data points to cluster |
%| |
%| varargin:  if run from concDt, varargin{1} = handle of "concDt" figure |
%| if run without GUI, varargin{1} = structure containing |
%| parameters store in the the substructure "param":  |
%| - param.xLim:  1x2 vector containing the limits of the x-axis |
%| - param.yLim:  1x2 vector containing the limits of the y-axis |
%| - param.maxIter:  maximum number of k-mean iterations |
%| |
%| - OUTPUT ------------------------------------------ |
%| |
%| means_cvg:  nx2 matrix containing the converged (x,y) coordinates of the n |
%| centres |
%| |
%| clusters:  1xn cell array containing the (x,y) coordinates of the data |
%| points belonging to each of the n cluster.  |
%| |
%| I_clust:  1xn vector containing the fraction of the total intensity in |
%| each cluster |
%| |
% Structure h contains the substructure "param"
if ~isstruct(varargin{1})
     h_fig = varargin{1};
```

```
    h = guidata(h_fig);
else
    h = varargin{1};
end
m = 0; % Nb of k-mean iteration
ok = 0;
means_cvg = means_start;
bin_x = (h.param.xLim(2) - h.param.xLim(1)) / h.param.xbin;
intv_x = h.param.xLim(1):bin_x:h.param.xLim(2);
bin_y = (h.param.yLim(2) - h.param.yLim(1)) / h.param.ybin;
intv_y = h.param.yLim(1):bin_y:h.param.yLim(2);
nbMeans = size(means_start,1); % Nb of clusters
% Find all data points with a non zero intensity
[id_y,id_x,I] = find(data);
% Get the corresponding x,y coordinates
y = mean([intv_y(id_y’); intv_y(id_y’+1)]);
x = mean([intv_x(id_x’); intv_x(id_x’+1)]);
xy = [x;y];
while ~ok && m <= h.param.maxIter
    % Initialise data points parameters for clustering
    excl = zeros(nbMeans,size(I,1));
    % Logical array:  exclude points from clusters
    dist = nan(nbMeans,size(I,1)); % Distances from the centres
    w_norm = zeros(nbMeans,size(I,1)); % Weights
    for i = 1:nbMeans
        Find data points lying in the tolerance zone of the cluster (circle)
        [o,c,o] = find(xy(2,:)  < (sqrt((tol(i,1)2) - (xy(1,:)  - ...
            means_cvg(i,1)).2) + means_cvg(i,2)) & ...
            xy(2,:)  > (-sqrt((tol(i,1)2) - (xy(1,:)  - ...
            means_cvg(i,1)).2) + means_cvg(i,2)));
        % Calculation of mean squared distances from centres (a,b)
        % dist = sqrt((a - x)2 + (b - y)2);
        dist(i,c) = sqrt(((means_cvg(i,1) - xy(1,c)).2) +
            ((means_cvg(i,2) - xy(2,c)).2));
        % Calculation of data points weights in the cluster
        % w = I;
        w_norm(i,c) = I(c,1)/sum(I(c,1),1); % Normalised weights
    end
    [o,id_mean] = min(dist,[],1);
    % Cluster where data points have the minimum distance
    excl = ~sum(w_norm,1);
    % Exclude all data points that belong to no cluster
    id_mean(:,excl) = 0;
    eq_clusters = 1;
    for i = 1:nbMeans
        % Cluster data points around each centre and calculate the coordinates
        % of the new centres
        [o,pnt,o] = find(id_mean == i);
        clusters{i} = [x(pnt)’, y(pnt)’];
        means_cvg(i,1) = sum((x(pnt)’.*(w_norm(i,pnt)/sum(w_norm(i,pnt),2))’),1);
```

```
          means_cvg(i,2) = sum((y(pnt)’.*(w_norm(i,pnt)/sum(w_norm(i,pnt),2))’),1);
          % Stop the iteration if the same data points still clustered together
          if m > 0 && prod(double(size(clusters{i}) ==
               size(clusters_prev{i})),2)
               eq_clusters = eq_clusters*prod(prod(double(clusters{i} ==
                    clusters_prev{i}),2),1);
          else
               eq_clusters = 0;
          end
     end
     % Stop k-mean process if the sum of the weights in each cluster doesn’t
     % change or if the max.  nb of iteration is reached
     if m == h.param.maxIter || (m > 0 && eq_clusters)
          ok = 1;
     else
          clusters_prev = clusters;
          m = m+1;
     end
end
% Calculate the fraction of the total intensity contained in each cluster
for n = 1:size(means_cvg,1)
     [o,indX,o] = find(intv_x <= means_cvg(n,1));
     means_cvg(n,1) = mean([intv_x(indX(numel(indX))) intv_x(indX(numel(indX))+1)]);
     [o,indY,o] = find(intv_y <= means_cvg(n,2));
     means_cvg(n,2) = mean([intv_y(indY(numel(indY))) intv_y(indY(numel(indY))+1)]);
     I_clust(n) = 0;
     for i = 1:size(clusters{n},1)
          [o,indX,o] = find(intv_x <= clusters{n}(i,2));
          [o,indY,o] = find(intv_y <= clusters{n}(i,1));
          I_clust(n) = I_clust(n) + data(indY(size(indY,2)), indX(size(indX,2)));
     end
end
if exist(’h_fig’, ’var’)
     setContPan([’Succes:  Center convergeance after ’ num2str(m) ’ cycles !’], ...
          ’success’, h_fig);
else
     disp([’Succes:  Center convergeance after ’ num2str(m) ’ cycles !’]);
end
```

### 12.3.4 loading_bar (written bar Mélodie Hadzic)

```
function stop = loading_bar(varargin)
% if nargin = 2 -> varagin{1} = status :  ’update’ or ’close’
% varagin{2} = h:  contains h.barData
%
% if nargin = 4 -> varagin{1} = status :  ’init’
% varagin{2} = h:  contains h.barData
% varagin{3} = nb_slopes:  the total number of steps
stop = 0;
```

```
h_bar = 24;
w_bar = 255;
img_bar = uint8(zeros(h_bar, w_bar, 3));
if nargin >= 2
     status = varargin{1};
     h_fig = varargin{2};
     h = guidata(h_fig);
     switch status
          case 'init'
               if nargin >= 3
                    if ~isfield(h, 'barData')
                    % register parameters in h
                    h.barData.curr_var = 0;
                    h.barData.nb_slopes = varargin{3};
                    % open the figure of the loading bar
                    h.barData.fig_main = gcf;
                    set(h.barData.fig_main, 'NextPlot', 'new');
                    h.barData.fig_bar = figure('Visible', 'off', 'MenuBar', ...
                         'none', 'Resize', 'off', 'Name', 'Process is running,
                         please wait...', 'NumberTitle', 'off', 'Units', 'pixels');
                    set(h.barData.fig_main, 'NextPlot', 'add');
                    % adjust dimensions of the new window
                    set(0, 'CurrentFigure', h.barData.fig_bar);
                    pos = get(h.barData.fig_bar, 'Position');
                    pos(3) = pos(3) * 2/3;
                    pos(4) = pos(4) / 4;
                    set(h.barData.fig_bar, 'Position', pos);
                    if nargin == 4
                         % create informative text
                         textInfo_bar = sprintf(varargin{4});
                         h.barData.textInfo_bar = uicontrol('Style', 'text', ...
                              'String', textInfo_bar, 'HorizontalAlignment', ...
                              'center', 'BackgroundColor', get(h.barData.fig_bar, ...
                              'Color'), 'ForegroundColor', 'black', 'Units', ...
                              'pixels','FontSize', 10);
                    end
                    % create progress text
                    text_bar = sprintf('Progress:  ');
                    h.barData.text_bar = uicontrol('Style', 'text', 'String', ...
                         text_bar, 'HorizontalAlignment', 'center', ...
                         'BackgroundColor', get(h.barData.fig_bar, ...
                         'Color'), 'Units', 'pixels');
                    % create image axes and show image bar
                    axes;
                    h.barData.img_bar = image(img_bar);
                    axis off;
                    axis image;
                    if nargin == 4
                         % set info text dimension and position
                         pos_textInfo = get(h.barData.textInfo_bar, 'Position');
```

```
                pos_textInfo(1) = 0;
                pos_textInfo(2) = pos(4) - 1.3*pos_textInfo(4);
                pos_textInfo(3) = pos(3);
                set(h.barData.textInfo_bar, 'Position', pos_textInfo);
            end
            % set progress text dimension and position
            pos_text = get(h.barData.text_bar, 'Position');
            pos_text(3) = pos(3);
            pos_text(1) = pos(3)/2 - pos_text(3)/2;
            set(h.barData.text_bar, 'Position', pos_text);
            % make the figure visible
            set(h.barData.fig_bar, 'Visible', 'on', 'CloseRequestFcn', ...
                {@fig_bar_CloseRequestFcn, h_fig});
        else
            h = rmfield(h, 'barData');
            guidata(h_fig, h);
            if nargin == 3
                loading_bar(varargin{1}, varargin{2}, varargin{3});
            else
                loading_bar(varargin{1}, varargin{2}, varargin{3}, ...
                    varargin{4});
            end
            h = guidata(h_fig);
        end
        else
            str = {'The number of input arguments is not correct' , ...
                'The status "init" requires 3 arguments'};
            stop = 1;
        end
    case 'update'
        if (isfield(h, 'barData') && isfield(h.barData, 'img_bar') && ...
            ishandle(h.barData.img_bar))
            if nargin == 2
                h.barData.curr_var = h.barData.curr_var + 1;
                if h.barData.curr_var >= (h.barData.prev_var + ...
                        h.barData.nb_slopes/200)
                    progress = h.barData.curr_var / h.barData.nb_slopes;
                    if progress <= 1
                        cursor_bar = progress * w_bar;
                        cursor_bar = uint8(cursor_bar);
                        % update the loading bar
                        for gg = 3:(cursor_bar - 2)
                            img_bar(3:(h_bar - 2), gg, 2) = 255;
                        end
                        set(h.barData.img_bar, 'CData', img_bar);
                        text_bar = sprintf('Progress:  %', progress*100);
                        set(h.barData.text_bar, 'String', text_bar);
                        h.barData.prev_var = h.barData.curr_var;
                        drawnow;
                    else
```

```
                                   close(h.barData.fig_bar);
                                   h = rmfield(h, ’barData’);
                                   str = {’The maximum number of loading steps
                                        reached.’,’Loading bar cannot be updated.’};
                                   stop = 1;
                              end
                         end
                    else
                         close(h.barData.fig_bar);
                         h = rmfield(h, ’barData’);
                         str = {’The number of imput arguments is not correct’ , ...
                              ’The status "update" requires 2 arguments’};
                         stop = 1;
                    end
               else
                    if isfield(h, ’barData’)
                         h = rmfield(h, ’barData’);
                    end
                    str = ’The loading bar does not exist:  it can not be updated.’;
                    stop = 1;
               end
          case ’close’
               if isfield(h, ’barData’) && ishandle(h.barData.fig_bar)
                    close(h.barData.fig_bar);
                    h = rmfield(h, ’barData’);
          end
     end
     guidata(h_fig, h);
     if stop == 1
          updateActPan(’Processed interrupted.’, h_fig);
          msgbox(str);
     end
else
     stop = 1;
     str = {’The number of imput arguments is not correct!’  , ’The function
          "loading_bar" requires at least 2 arguments.’};
     msgbox(str);
end
function fig_bar_CloseRequestFcn(hObj, evd, h_fig)
h = guidata(h_fig);
if isfield(h, ’barData’)
     h = rmfield(h, ’barData’);
     guidata(h_fig, h);
end
delete(hObj);
```

### 12.3.5 mmexpfit

```
function [fitres,ex_, cf_, gof] = mmexpfit(x_data,y_data,fitparam,fit_type_nr)
```

```
% redefine fit parameters depending on user selection
% y-offset
if fitparam.lower(1,1) == 0 && fitparam.start(1,1) == 0 && fitparam.upper(1,1) == 0
    y_offset = 0;
else
    y_offset = 1;
end
% x-offset
if fitparam.lower(1,2) == 0 && fitparam.start(1,2) == 0 && fitparam.upper(1,2) == 0
    x_offset = 0;
else
    x_offset = 1;
end
% rescale fitparameter arrays
if y_offset == 0 && x_offset == 0
    fitparam.lower = fitparam.lower(1,3:size(fitparam.lower,2));
    fitparam.start = fitparam.start(1,3:size(fitparam.start,2));
    fitparam.upper = fitparam.upper(1,3:size(fitparam.upper,2));
elseif y_offset == 0 && x_offset ~= 0
    fitparam.lower = fitparam.lower(1,2:size(fitparam.lower,2));
    fitparam.start = fitparam.start(1,2:size(fitparam.start,2));
    fitparam.upper = fitparam.upper(1,2:size(fitparam.upper,2));
elseif y_offset ~= 0 && x_offset == 0
    fitparam.lower = [fitparam.lower(1,1), ...
        fitparam.lower(1,3:size(fitparam.lower,2))];
    fitparam.start = [fitparam.start(1,1), ...
        fitparam.start(1,3:size(fitparam.start,2))];
    fitparam.upper = [fitparam.upper(1,1), ...
        fitparam.upper(1,3:size(fitparam.upper,2))];
end
% Define fit type
switch fit_type_nr
    case 1
        % Single-exponential decay
        if y_offset == 0 && x_offset == 0
            ft_ = fittype('offset+a*exp(-(x-x0)/b)',...
                'dependent',{'y'},'independent',{'x'},...
                'problem', {'offset', 'x0'}, ...
                'coefficients',{'a', 'b'});
        elseif y_offset ~= 0 && x_offset == 0
            ft_ = fittype('offset+a*exp(-(x-x0)/b)',...
                'dependent',{'y'},'independent',{'x'},...
                'problem', 'x0', ...
                'coefficients',{'offset', 'a', 'b'});
        elseif y_offset == 0 && x_offset ~= 0
            ft_ = fittype('offset+a*exp(-(x-x0)/b)',...
                'dependent',{'y'},'independent',{'x'},...
                'problem', 'offset', ...
                'coefficients',{'x0', 'a', 'b'});
        else
```

```
               ft_ = fittype(’offset+a∗exp(-(x-x0)/b)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’coefficients’,{’offset’, ’x0’, ’a’, ’b’});
          end
     case 2
          % Bi-exponential decay
          fitparam.lower(size(fitparam.lower,2)-1) = [];
          fitparam.start(size(fitparam.start,2)-1) = [];
          fitparam.upper(size(fitparam.upper,2)-1) = [];
          if y_offset == 0 && x_offset == 0
               ft_ = fittype(’offset+a∗exp(-(x-x0)/b)+(1-a)∗exp(-(x-x0)/d)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, {’offset’, ’x0’}, ...
                    ’coefficients’,{’a’, ’b’, ’d’});
          elseif y_offset ∼= 0 && x_offset == 0
               ft_ = fittype(’offset+a∗exp(-(x-x0)/b)+(1-a)∗exp(-(x-x0)/d)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, ’x0’, ...
                    ’coefficients’,{’offset’, ’a’, ’b’, ’d’});
          elseif y_offset == 0 && x_offset ∼= 0
               ft_ = fittype(’offset+a∗exp(-(x-x0)/b)+(1-a)∗exp(-(x-x0)/d)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, ’offset’, ...
                    ’coefficients’,{’x0’, ’a’, ’b’, ’d’});
          else
               ft_ = fittype(’offset+a∗exp(-(x-x0)/b)+(1-a)∗exp(-(x-x0)/d)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’coefficients’,{ ’offset’, ’x0’, ’a’, ’b’, ’d’});
          end
     case 3
          % Tri-exponential decay
          fitparam.lower(size(fitparam.lower,2)-1) = [];
          fitparam.start(size(fitparam.start,2)-1) = [];
          fitparam.upper(size(fitparam.upper,2)-1) = [];
          if y_offset == 0 && x_offset == 0
               ft_ = fittype(’offset+a∗exp(-(x-x0)/b)+ ...
                    c∗exp(-(x-x0)/d)+(1-a-c)∗exp(-(x-x0)/f)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, {’offset’, ’x0’}, ...
                    ’coefficients’,{’a’, ’b’, ’c’, ’d’, ’f’});
          elseif y_offset ∼= 0 && x_offset == 0
               ft_ = fittype(’offset+a∗exp(-(x-x0)/b)+ ...
                    c∗exp(-(x-x0)/d)+(1-a-c)∗exp(-(x-x0)/f)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, ’x0’, ...
                    ’coefficients’,{’offset’, ’a’, ’b’, ’c’, ’d’, ’f’});
          elseif y_offset == 0 && x_offset ∼= 0
               ft_ = fittype(’offset+a∗exp(-(x-x0)/b)+ ...
                    c∗exp(-(x-x0)/d)+(1-a-c)∗exp(-(x-x0)/f)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
```

```
                    ’problem’, ’offset’, ...
                    ’coefficients’,{’x0’, ’a’, ’b’, ’c’, ’d’, ’f’});
          else
               ft_ = fittype(’offset+a∗exp(−(x−x0)/b)+ ...
                    c∗exp(−(x−x0)/d)+(1−a−c)∗exp(−(x−x0)/f)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’coefficients’,{ ’offset’, ’x0’, ’a’, ’b’, ’c’, ’d’, ’f’});
          end
     case 4
          % Tetra−exponential decay
          fitparam.lower(size(fitparam.lower,2)−1) = [];
          fitparam.start(size(fitparam.start,2)−1) = [];
          fitparam.upper(size(fitparam.upper,2)−1) = [];
          if y_offset == 0 && x_offset == 0
               ft_ = fittype(’offset+a∗exp(−(x−x0)/b)+c∗exp(−(x−x0)/d)+ ...
                    e∗exp(−(x−x0)/f)+(1−a−c−e)∗exp(−(x−x0)/h)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, {’offset’, ’x0’}, ...
                    ’coefficients’,{’a’, ’b’, ’c’, ’d’, ’e’, ’f’, ’h’});
          elseif y_offset ∼= 0 && x_offset == 0
               ft_ = fittype(’offset+a∗exp(−(x−x0)/b)+c∗exp(−(x−x0)/d)+ ...
                    e∗exp(−(x−x0)/f)+(1−a−c−e)∗exp(−(x−x0)/h)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, ’x0’, ...
                    ’coefficients’,{’offset’, ’a’, ’b’, ’c’, ’d’, ’e’, ’f’, ’h’});
          elseif y_offset == 0 && x_offset ∼= 0
               ft_ = fittype(’offset+a∗exp(−(x−x0)/b)+c∗exp(−(x−x0)/d)+ ...
                    e∗exp(−(x−x0)/f)+(1−a−c−e)∗exp(−(x−x0)/h)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, ’offset’, ...
                    ’coefficients’,{’x0’, ’a’, ’b’, ’c’, ’d’, ’e’, ’f’, ’h’});
          else
               ft_ = fittype(’offset+a∗exp(−(x−x0)/b)+c∗exp(−(x−x0)/d)+ ...
                    e∗exp(−(x−x0)/f)+(1−a−c−e)∗exp(−(x−x0)/h)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’coefficients’,{ ’offset’, ’x0’, ’a’, ’b’, ’c’, ’d’, ...
                         ’e’, ’f’, ’h’});
          end
     case 5
          % Stretched exponential decay
          if y_offset == 0 && x_offset == 0
               ft_ = fittype(’offset+a∗exp(−((x−x0)/b)c)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, {’offset’, ’x0’}, ...
                    ’coefficients’,{’a’, ’b’, ’c’});
          elseif y_offset ∼= 0 && x_offset == 0
               ft_ = fittype(’offset+a∗exp(−((x−x0)/b)c)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, ’x0’, ...
                    ’coefficients’,{’offset’, ’a’, ’b’, ’c’});
```

```
          elseif y_offset == 0 && x_offset ∼= 0
               ft_ = fittype(’offset+a⋆exp(-((x-x0)/b)c)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’problem’, ’offset’, ...
                    ’coefficients’,{’x0’, ’a’, ’b’, ’c’});
          else
               ft_ = fittype(’offset+a⋆exp(-((x-x0)/b)c)’,...
                    ’dependent’,{’y’},’independent’,{’x’},...
                    ’coefficients’,{ ’offset’, ’x0’, ’a’, ’b’, ’c’});
          end
end
% Set fit options
excl_x = [x_data(1) x_data(size(x_data,1))];
ex_ = logical(zeros(length(x_data),1));
ex_([[]]) = 1;
ex_ = ex_ | ∼(excl_x(1)<x_data & x_data<excl_x(2));
ok_ = ∼(isnan(x_data) | isnan(y_data));
st_ = fitparam.start;
fo_ = fitoptions(’method’,’NonlinearLeastSquares’,’Lower’,fitparam.lower,...
     ’Upper’,fitparam.upper,’Startpoint’,st_,’Exclude’,ex_(ok_),’MaxIter’, 1000);
% Fit model to data
if sum(∼ex_(ok_))<2
% too many points excluded
     error(’Not enough data left to fit’
else
     if y_offset == 0 && x_offset == 0
          [cf_, gof, output] = fit(x_data(ok_),y_data(ok_),ft_,fo_, ...
               ’problem’, {0, 0});
     elseif y_offset == 0 || x_offset == 0
          [cf_, gof, output] = fit(x_data(ok_),y_data(ok_),ft_,fo_, ...
               ’problem’, {0});
     else
          [cf_, gof, output] = fit(x_data(ok_),y_data(ok_),ft_,fo_);
     end
     % get fitting results, reformat
     coeffs = coeffvalues(cf_);
     if x_offset == 0 && y_offset == 0
          coeffs = [0, 0, coeffs];
     elseif x_offset ∼= 0 && y_offset == 0
          coeffs = [0, coeffs];
     elseif x_offset == 0 && y_offset ∼= 0
          coeffs = [coeffs(1,1), 0, coeffs(1,2:size(coeffs,2))];
     end
     fitres.refxo = coeffs(1,2);
     fitres.refyo = coeffs(1,1);
     if fit_type_nr == 5
          fitres.refa = coeffs(1,3);
          fitres.refc = coeffs(1,4);
          fitres.refbeta = coeffs(1,5);
     else
```

```
        if fit_type_nr == 2
            coeffs(1,size(coeffs,2):(size(coeffs,2)+1)) = [1-coeffs(1,3),...
                coeffs(1,size(coeffs,2))];
        elseif fit_type_nr == 3
            coeffs(1,size(coeffs,2):(size(coeffs,2)+1)) = [(1-coeffs(1,3)- ...
                coeffs(1,5)),coeffs(1,size(coeffs,2))];
        elseif fit_type_nr == 4
            coeffs(1,size(coeffs,2):(size(coeffs,2)+1)) = [(1-coeffs(1,3)- ...
                coeffs(1,5)-coeffs(1,7)),coeffs(1,size(coeffs,2))];
        end
        for i = 1:fit_type_nr
            fitres.refa(1,i) = coeffs(1,(3+(i-1)*2));
            fitres.refc(1,i) = coeffs(1,(4+(i-1)*2));
        end
    end
end
```

### 12.3.6 mmgaussfit.m

```
function [fitres, aic, ex_, cf_, gof] = mmgaussfit(x_data,y_data,fitparam,...
    fit_type_nr,weight)
% reformat input data
ex_ = logical(zeros(length(y_data),1));
ex_([[]]) = 1;
ex_ = ex_ | ~(x_data(1)<x_data &x_data< x_data(size(x_data,1)));
    for j=2:fit_type_nr
        % define fit options
        fo_ = fitoptions('method','NonlinearLeastSquares', ...
            'Lower',fitparam.lower(1,1:(3*j)),'Upper',fitparam.upper(1,1:(3*j)));
        ok_ = ~(isnan(x_data) | isnan(y_data));
        st_ = fitparam.start(1,1:(3*j));
        set(fo_, 'Startpoint', st_, 'Exclude', ex_(ok_));
        % weighting is generally disabled in BOBA FRET (weight = 0), but
        % can in principle be activated
        if weight > 0
            set(fo_,'Weights',x_data(ok_).weight);
        end
        str_fittype = ['gauss' num2str(j)];
        ft_ = fittype(str_fittype);
        if sum(~ex_(ok_))<2
        % too many points excluded
            error('Not enough data left to fit "%s" after applying exclusion
                rule "%s".','fit 1','Exlc_D');
        else
            [cf_, gof, output] = fit(x_data(ok_),y_data(ok_),ft_,fo_);
        end
        coeffs = coeffvalues(cf_);
        fitres.Adj_S = gof.adjrsquare;
        for gg=1:j
```

```
            fitres.refa(1,gg) = coeffs(1,3*gg-2); % Gaussparameter(s) amplitude
            fitres.refb(1,gg) = coeffs(1,3*gg-1); % Gaussparameter(s) center
            fitres.refc(1,gg) = coeffs(1,3*gg-0); % Gaussparameter(s) width
        end
        % An AIC (Akaike information criterion) was originally implemented
        % for model selection, but it turned out to be too selective.  Here,
        % "aic" corresponds to the root mean squared error.
        aic(1,(j-1)) = gof.rmse;
    end
```

### 12.3.7 mmimportdata

```
function [ok, histref, filelist, histfiles, histarray, weighting_vector] = ...
    mmimportdata(thAna, histDat, h_fig, overflowbins)
h = guidata(h_fig);
ok = 0;
histref = [];
filelist = {};
histfiles = [];
histarray = [];
folder = get(h.edit_folder, 'String');
filePattern = fullfile(folder, ['*.'  get(h.edit_filetype, 'String')]);
% import files
histfiles = dir(filePattern);
if ~isdir(folder)
    errorMessage = sprintf('Error:  The following folder does not exist:\n%s', ...
        folder);
    uiwait(warndlg(errorMessage));
    return;
elseif size(histfiles,1) == 0
    errorMessage = sprintf('Error:  No files found.');
    uiwait(warndlg(errorMessage));
    return;
end
if thAna && histDat
    % thermodynamic analysis from FRET histograms
    weighting = get(h.radiobutton_weighing_yes, 'Value');
    for k = 1:length(histfiles)
        fullFileName = fullfile(folder, histfiles(k).name);
        histarray{k} = dlmread(fullFileName);
        % check whether matrix dimensions agree
        checkdims = isequal(size(histarray{1}), size(histarray{k}));
        if checkdims == 0
            errorMessage = 'Error:  Matrix dimensions are inconsistent';
            uiwait(warndlg(errorMessage));
            return;
        end
        % create list of histogram files
        filelist{k,1} = histfiles(k,1).name;
```

```
     end
     if weighting
          % histograms are weighted regarding the number of data point
          histref = zeros(size(histarray{1}));
          for k = 1:size(histarray,2)
               histref = histref + histarray{1,k};
          end
          if overflowbins == 0
               % remove first & last FRET bin
               histref(size(histref,1),:)  = [];
               histref(1,:)  = [];
          end
          histref(:,1) = histref(:,1)/size(histarray,2);
          % restore FRET values
          histref(:,size(histref,2)) = histref(:,size(histref,2))/ ...
               histref(size(histref,1),size(histref,2));
          % re-normalize 1-cumP
          histref(:,2) = histref(:,2)/sum(histref(:,2));
          % restore occurences
     else
          % histograms have the same weight
          histref = zeros(size(histarray{1},1),size(histarray{1},2));
          if overflowbins == 0
               histref = histref((1:size(histref,1)-2),:);
          end
          for k = 1:size(histarray,2)
               if overflowbins == 0
                    histarray{1,k}(size(histarray{1,k},1),:)  = [];
                    histarray{1,k}(1,:)  = [];
               end
               histarray{1,k}(:,2) = histarray{1,k}(:,2)/sum(histarray{1,k}(:,2));
               histref = histref + histarray{1,k};
          end
          histref(:,1) = histref(:,1)/size(histarray,2);
          histref(:,2) = histref(:,2)/size(histarray,2);
          % not relevant, solely first two columns used.
          % histref(:,3) = histref(:,2);
          % histref(:,size(histref,2)) = 1 - histref(:,size(histref,2))/ ...
               % histref(size(histref,1),size(histref,2));
          % histref(:,4) = 1 - histref(:,size(histref,2));
     end
     set(h.display_results, ’UserData’, []);
     weighting_vector = ones(size(histfiles,1),1);
     for i = 1:size(histfiles,1)
          weighting_vector(i,1) = sum(histarray{1,i}(:,2));
     end
     weighting_vector = weighting_vector/sum(weighting_vector);
else
     % thermodynamic or kinetic analysis using dwelltime files
     % get transitions
```

```
if isempty(get(h.display_results, 'UserData'))
    str_quest = 'Do you want to cluster the transitions?';
else
    str_quest = 'Do you want to re-cluster the transitions?';
end
choice1 = questdlg(str_quest, 'Transition clustering', 'Yes', 'No', 'Yes');
choice2 = questdlg('The first and the last dwell time are usually not reliably
        determined.  Would you like to discard them?', ...
    'Discard dwell times', 'Yes', 'No', 'Yes');
if strcmp(choice1, 'No') && ~isempty(get(h.display_results, 'UserData'))
    dat = get(h.display_results, 'UserData');
    filelist = dat.filelist;
    histfiles = dat.histfiles;
    histarray = dat.histarray;
    if strcmp(choice2, 'Yes')
        indexing = [];
        for k = 1:size(histarray,2)
            if (size(histarray{k},1)<3)
                indexing = [indexing, k];
            end
        end
        histarray(indexing) = [];
        filelist(indexing) = [];
        histfiles(indexing) = [];
        for k = 1:size(histarray,2)
            histarray{1,k}(1,:)  = [];
            histarray{1,k}(size(histarray{1,k},1),:)  = [];
        end
    end
else
    histarray = [];
    for k = 1:length(histfiles)
        fullFileName = fullfile(folder, histfiles(k).name);
        histarray{k} = dlmread(fullFileName);
        checkdims = isequal(size(histarray{1},2), size(histarray{k},2));
        if checkdims == 0
            errorMessage = 'Error:  Matrix dimensions are inconsistent';
            uiwait(warndlg(errorMessage));
            return;
        end
        filelist{k,1} = histfiles(k,1).name;
    end
    if strcmp(choice2, 'Yes')
        indexing = [];
        for k = 1:size(histarray,2)
            if (size(histarray{k},1)<3)
                indexing = [indexing, k];
            end
        end
        histarray(indexing) = [];
```

```
            filelist(indexing) = [];
            histfiles(indexing) = [];
            for k = 1:size(histarray,2)
                histarray{1,k}(1,:)  = [];
                histarray{1,k}(size(histarray{1,k},1),:)  = [];
            end
        end
        if strcmp(choice1, 'Yes') % cluster the transitions
            histarray = fCell2fMat(histarray);
            h_fig_clustTrans = clustTrans(histarray);
            set(h_fig, 'Visible', 'off');
            uiwait(h_fig_clustTrans);
            set(h_fig, 'Visible', 'on');
            m = guidata(h_fig_clustTrans);
            histarray = m.histarray;
            histarray = fMat2fCell(histarray);
            if ishandle(h_fig_clustTrans)
                close(h_fig_clustTrans);
            end
            if isempty(histarray)
                return;
            end
        end
    end
    dat.filelist = filelist;
    dat.histfiles = histfiles;
    dat.histarray = histarray;
    set(h.display_results, 'UserData', dat);
    transitions = [];
    indexing = [];
    for k = 1:size(histarray,2)
        if isempty(find(isnan(histarray{k}))) == 1
            transitions = [transitions; histarray{k}];
        else
            indexing(1,(size(indexing,2)+1)) = k;
        end
    end
    transitions = unique(transitions(:,(size(transitions,2)- ...
        1):size(transitions,2)), 'rows');
    histarray(indexing) = [];
    weighting_vector = ones(size(histarray,2),1);
    for i = 1:size(histarray,2)
        weighting_vector(i,1) = sum(histarray{1,i}(:,1));
    end
    if ~thAna % kinetic analysis
        % Construct a questdlg to select the FRET transition to be analysed
        for k = 1:size(transitions,1)
            transstr{k,1} = num2str(transitions(k,:));
            prompt{k,:} = strcat(num2str(transitions(k,1)),' => ', ...
                num2str(transitions(k,2)));
```

```
          end
          [index,ok2] = listdlg(’PromptString’,’Select a FRET transition:’,...
               ’SelectionMode’,’single’,...
               ’ListString’,prompt);
          % continue if a selection has been made, otherwise terminate.
          if ok2
               % select and sort dwell times
               for k = 1:size(histarray,2)
                    indexing = ismember(histarray{1,k}(:,(size(histarray{1,k}, ...
                         2)-1):size(histarray{1,k}, ...
                         2)),transitions(index,:),’rows’);
                    histarray{1,k} = histarray{1,k}(indexing,:);
               end
               indexing = [];
               for k = 1:size(histarray,2)
                    if (size(histarray{k},1)<1)
                         indexing = [indexing, k];
                    end
               end
               histarray(indexing) = [];
               weighting_vector(indexing) = [];
               weighting = get(h.radiobutton_weighing_yes, ’Value’);
               if weighting
                    weighting_vector = weighting_vector/sum(weighting_vector);
               else
                    weighting_vector = ones(size(weighting_vector,1),1);
               end
               histref = [];
               for k = 1:size(histarray,2)
                    histref(1:(size(histref,1)+size(histarray{k},1)),1:2) = ...
                         [histref; histarray{k}(:,1), ...
                         zeros(size(histarray{k},1),1)];
                    histref(((size(histref,1)-size(histarray{k}(:,1),1)+ ...
                         1):size(histref,1)),2) = weighting_vector(k);
               end
               histref = sortrows(histref,1);
               histref(:,2) = 1 - linspace(1, size(histref,1), ...
                    size(histref,1))/size(histref,1);
               [C,ia,ic] = unique(histref(:,1));
               histref = histref(ia,:);
               histref = [0, 1; histref]; % flipud
          else
               return;
          end
     else % thermodynamic analysis using dwell times
          histref = zeros(size(transitions,1),3);
          histref(:,2:3) = transitions;
          weighting = get(h.radiobutton_weighing_yes, ’Value’);
          if weighting
               weighting_vector = weighting_vector/sum(weighting_vector);
```

```
     for i = 1:size(histarray,2)
          histarray{1,i} = sortrows(histarray{1,i},2);
          [C,ia,ic] = unique(histarray{1,i}(:,2));
          for k = 1:size(ia,1)
               if k == 1
                    histarray_norm(k,1:3) = [sum(histarray{1,i}(1:ia(k,...
                         1),1)), histarray{1,i}(ia(k,1),2:3)];
               else
                    histarray_norm(k,1:3) = [sum(histarray{1,i}(ia((k-1),...
                         1):ia(k,1),1)), histarray{1, ...
                         i}(ia(k,1),2:3)];
               end
          end
          if size(histarray_norm) == size(histref)
               histref(:,1) = histref(:,1) + histarray_norm(:,1);
          else
               for l = 1:size(histarray_norm,1)
                    j = 1;
                    while isequal(histarray_norm(l,2), histref(j,2)) == 0
                         j = j + 1;
                    end
                    histref(j,1) = histref(j,1) + histarray_norm(l,1);
               end
          end
          histarray{1,i} = histarray_norm;
          histarray{1,i}(:,1) = histarray{1,i}(:,1)/ ...
               sum(histarray{1,i}(:,1));
     end
else
     weighting_vector = ones(size(weighting_vector,1),1);
     for i = 1:size(histarray,2)
          histarray{1,i} = sortrows(histarray{1,i},2);
          [C,ia,ic] = unique(histarray{1,i}(:,2));
          for k = 1:size(ia,1)
               if k == 1
                    histarray_norm(k,1:3) = [sum(histarray{1, ...
                         i}(1:ia(k,1),1)), histarray{1,i}(ia(k,1),2:3)];
               else
                    histarray_norm(k,1:3) = [sum(histarray{1, ...
                         i}(ia((k-1),1):ia(k,1),1)), histarray{1, ...
                         i}(ia(k,1),2:3)];
               end
          end
          histarray{1,i} = histarray_norm;
          histarray{1,i}(:,1) = histarray{1,i}(:,1)/sum(histarray{1, ...
               i}(:,1));
          if size(histarray{1,i}) == size(histref)
               histref(:,1) = histref(:,1) + histarray{1,i}(:,1);
          else
               for l = 1:size(histarray{1,i},1)
```

```
                        j = 1;
                        while isequal(histarray{1,i}(1,2), histref(j,2)) == 0
                            j = j + 1;
                        end
                        histref(j,1) = histref(j,1) + histarray{1,i}(1,1);
                    end
                end
            end
        end
        histref(:,1) = histref(:,1)/sum(histref(:,1));
    end
end
ok = 1;

function hist_mat = fCell2fMat(hist_cell)
hist_mat = zeros(1, size(hist_cell{1},2)+1);
for i = 1:size(hist_cell,2)
    hist_mat = [hist_mat; [hist_cell{i} ones(size(hist_cell{i},1),1)*i]];
end
hist_mat = hist_mat(2:size(hist_mat,1), :);

function hist_cell = fMat2fCell(hist_mat)
n_file = max(hist_mat(:,size(hist_mat,2)));
hist_cell = cell(1,n_file);
for i = 1:n_file
    hist_cell{i} = hist_mat((hist_mat(:,size(hist_mat,2)) == i), ...
        1:size(hist_mat,2) - 1);
end
```

### 12.3.8 mmsavedata

```
function mmsavedata(param, res, h_axes)
thAna = param(1); % (thermodynamic/kinetic) analysis
histDat = param(2); % from (histogram/dwell-times) data
threshAna = param(3); % by (thresholding/gauss fit) method
filename_root = ’results_randomisation’;
if thAna && histDat && ∼threshAna
    % thermodynamic analysis using histogram files by gauss fit
    % files to save
    n = getNbFiles(filename_root, {’_amplitude.txt’, ’_center.txt’, ...
        ’_width.txt’, ’_reloccurrence.txt’, ’_plot.jpg’, ’_plot.fig’});
    str_n = ”;
    if n > 1
        str_n = [’(’ num2str(n) ’)’];
    end
    filename_amp = [filename_root str_n ’_amplitude.txt’];
    filename_ctr = [filename_root str_n ’_center.txt’];
    filename_w = [filename_root str_n ’_width.txt’];
    filename_relocc = [filename_root str_n ’_reloccurrence.txt’];
```

```
% results to write
amp_ref = res.fitres2.refa;
amp = res.fitres2.amplitudessorted;
amp_mean = mean(amp);
amp_std = std(amp);
ctr_ref = res.fitres2.refb;
ctr = res.fitres2.centersorted;
ctr_mean = mean(ctr);
ctr_std = std(ctr);
w_ref = res.fitres2.refc/sqrt(2);
w = res.fitres2.widthsorted/sqrt(2);
w_mean = mean(w);
w_std = std(w);
occ_ref = res.fitres2.reloccurrenceref;
occ = res.fitres2.reloccurrencesorted;
occ_mean = mean(occ);
occ_std = std(occ);
fmt = [];
for i = 1:size(amp_ref,2)
     fmt = [fmt ’%d’];
end
clear res;
% -----------
f_amp = fopen(filename_amp, ’Wt’);
fprintf(f_amp, [’Reference amplitudes\n’ fmt ’\n’ ...
          ’Bootstrap amplitudes (mean)\n’ fmt ’\n’ ...
          ’Bootstrap standard deviation (1*sigma)\n’ fmt ’\n’ ...
          ’Bootstrap amplitudes (sorted)\n’], ...
          amp_ref, amp_mean, amp_std);
fprintf(f_amp, [fmt ’\n’], amp);
fclose(f_amp);
f_ctr = fopen(filename_ctr, ’Wt’);
fprintf(f_ctr, [’Reference center values\n’ fmt ’\n’ ...
          ’Bootstrap center values (mean)\n’ fmt ’\n’ ...
          ’Bootstrap standard deviation (1*sigma)\n’ fmt ’\n’ ...
          ’Bootstrap center values (sorted)\n’], ...
          ctr_ref, ctr_mean, ctr_std);
fprintf(f_ctr, [fmt ’\n’], ctr);
fclose(f_ctr);
f_w = fopen(filename_w, ’Wt’);
fprintf(f_w, [’Reference width values\n’ fmt ’\n’ ...
          ’Bootstrap width values (mean)\n’ fmt ’\n’ ...
          ’Bootstrap standard deviation (1*sigma)\n’ fmt ’\n’ ...
          ’Bootstrap width values (sorted)\n’], ...
          w_ref, w_mean, w_std);
fprintf(f_w, [fmt ’\n’], w);
fclose(f_w);
f_relocc = fopen(filename_relocc, ’Wt’);
fprintf(f_relocc, [’Reference relative occurrences\n’ fmt ’\n’ ...
          ’Bootstrap relative occurrences (mean)\n’ fmt ’\n’ ...
```

```
               ’Bootstrap standard deviation (1∗sigma)\n’ fmt ’\n’ ...
               ’Bootstrap relative occurrences (sorted)\n’], ...
               occ_ref, occ_mean, occ_std);
     fprintf(f_relocc, [fmt ’\n’], occ);
     fclose(f_relocc);
elseif thAna && ((histDat && threshAna) || ∼histDat)
     % thermodynamic analysis using dwell-time files
     % or using histogram files by thresholding
     % file to save
     n = getNbFiles(filename_root, {’.txt’, ’_plot.jpg’, ’_plot.fig’});
     str_n = ”;
     if n > 1
          str_n = [’(’ num2str(n) ’)’];
     end
     filename = [filename_root str_n ’.txt’];
     % results to write
     fmt = [];
     states = [];
     for i = 1:size(res.states,2)
          states = [states res.states{1,i}];
          fmt = [fmt ’%d’];
     end
     rel_occ = res.reference;
     btsp_mean = mean(res.population);
     btsp_std = std(res.population);
     btsp_rel = transpose(res.population);
     % -----------
     f = fopen(filename, ’Wt’);
     fprintf(f, [’Population of the different FRET states\n%s\n’ ...
               ’Relative occurrence (reference, normalised to 1)\n’ fmt ’\n’ ...
               ’Mean relative occurrence (bootstrap, normalised to 1)\n’fmt’\n’...
               ’Bootstrap standard deviation (1∗sigma)\n’ fmt ’\n’ ...
               ’Bootstrap relative occurrences (sorted)\n’ fmt ’\n’], ...
               states, rel_occ, btsp_mean, btsp_std);
     fprintf(f, [fmt ’\n’], btsp_rel);
     fclose(f);
elseif ∼thAna % kinetic analysis
     fit_type_nr = res.fit_type_nr;
     % files to save
     n = getNbFiles(filename_root, {’_amplitude.txt’, ’_decconstants.txt’, ...
          ’_x-offset.txt’, ’_y-offset.txt’, ’_beta.txt’, ’_plot.jpg’, ...
          ’_plot.fig’});
     str_n = ”;
     if n > 1
          str_n = [’(’ num2str(n) ’)’];
     end
     filename_amp = [filename_root str_n ’_amplitude.txt’];
     filename_cst = [filename_root str_n ’_decconstants.txt’];
     filename_x_os = [filename_root str_n ’_x-offset.txt’];
     filename_y_os = [filename_root str_n ’_y-offset.txt’];
```

```
if fit_type_nr == 5
     filename_b = [filename_root str_n ’_beta.txt’];
end
% results to write
amp_ref = res.fitres2.refa;
cst_ref = res.fitres2.refc;
if fit_type_nr == 5
     amp = res.fitres2.sorted(:,5);
     cst = res.fitres2.sorted(:,2);
     beta_ref = res.fitres2.refbeta;
beta = res.fitres2.sorted(:,1);
     beta_mean = mean(beta);
     beta_std = std(beta);
else
     amp = res.fitres2.sorted(:,((fit_type_nr + 3):(fit_type_nr + ...
          3)+fit_type_nr-1));
     cst = res.fitres2.sorted(:,(1:fit_type_nr));
end
amp_mean = mean(amp);
amp_std = std(amp);
cst_mean = mean(cst);
cst_std = std(cst);
if res.x_offset
     x_ref = res.fitres2.refxo;
     if fit_type_nr == 5
          x = res.fitres2.sorted(:,4);
     else
          x = res.fitres2.sorted(:,(fit_type_nr + 2));
     end
     x_mean = mean(x);
     x_std = std(x);
end
if res.y_offset
     y_ref = res.fitres2.refyo;
     if fit_type_nr == 5
          y = res.fitres2.sorted(:,3);
     else
          y = res.fitres2.sorted(:,(fit_type_nr + 1));
     end
     y_mean = mean(y);
     y_std = std(y);
end
fmt = [];
for i = 1:size(amp_ref,2)
     fmt = [fmt ’%d’];
end
clear res;
% -----------
f_amp = fopen(filename_amp, ’Wt’);
fprintf(f_amp, [’Reference amplitudes\n’ fmt ’\n’ ...
```

```
               'Bootstrap amplitudes (mean)\n' fmt '\n' ...
               'Bootstrap standard deviation (1*sigma)\n' fmt '\n' ...
               'Bootstrap amplitudes (sorted)\n'], ...
               amp_ref, amp_mean, amp_std);
     fprintf(f_amp, [fmt '\n'], amp);
     fclose(f_amp);
     f_cst = fopen(filename_cst, 'Wt');
     fprintf(f_cst, ['Reference decay constants\n' fmt '\n' ...
               'Bootstrap decay constants (mean)\n' fmt '\n' ...
               'Bootstrap standard deviation (1*sigma)\n' fmt '\n' ...
               'Bootstrap decay constants (sorted)\n'], ...
               cst_ref, cst_mean, cst_std);
     fprintf(f_cst, [fmt '\n'], cst);
     fclose(f_cst);
     f_x = fopen(filename_x_os, 'Wt');
     if ~exist('x', 'var')
          fprintf(f_x, 'x-offset\n0\n');
     else
          fprintf(f_x, ['Reference x-offset\n' fmt '\n' ...
               'Bootstrap x-offsets (mean)\n' fmt '\n' ...
               'Bootstrap standard deviation (1*sigma)\n' fmt '\n' ...
               'Bootstrap x-offsets (sorted)\n'], ...
               x_ref, x_mean, x_std);
          fprintf(f_x, [fmt '\n'], x);
     end
     fclose(f_x);
     f_y = fopen(filename_y_os, 'Wt');
     if ~exist('y', 'var')
          fprintf(f_y, 'y-offset\n0\n');
     else
          fprintf(f_y, ['Reference y-offset\n' fmt '\n' ...
               'Bootstrap y-offsets (mean)\n' fmt '\n' ...
               'Bootstrap standard deviation (1*sigma)\n' fmt '\n' ...
               'Bootstrap y-offsets (sorted)\n'], ...
               y_ref, y_mean, y_std);
          fprintf(f_y, [fmt '\n'], y);
     end
     fclose(f_y);
     if fit_type_nr == 5
          f_b = fopen(filename_b, 'Wt');
          fprintf(f_b, ['Reference beta\n' fmt '\n' ...
               'Bootstrap betas (mean)\n' fmt '\n' ...
               'Bootstrap standard deviation (1*sigma)\n' fmt '\n' ...
               'Bootstrap betas (sorted)\n'], ...
               beta_ref, beta_mean, beta_std);
          fprintf(f_b, [fmt '\n'], beta);
          fclose(f_b);
     end
end
filename_plot = [filename_root str_n '_plot'];
```

```
h_p = figure('Visible', 'off', 'Color', [1 1 1], 'Units', get(h_axes, 'Units'));
h_a = copyobj(h_axes, h_p);
grid(h_a, 'on');
set(h_a, 'Position', [0.06 0.06 0.88 0.88]);
set(h_p, 'Position', get(h_axes, 'Position'));
saveas(h_p, filename_plot, 'jpg');
% saveas(h_p, filename_plot, 'fig');
close(h_p);

function n = getNbFiles(rootName, fNames);
n = 1;
p = what();
for i = 1:size(fNames,2)
    if exist([p.path '\' rootName fNames{i}], 'file')
        n_i = max([n;2]);
        [o,subName,ext] = fileparts(fNames{i});
        while exist([p.path '\' rootName '(' num2str(n_i) ')' subName ext], 'file')
            n_i = n_i + 1;
        end
        if n_i > n
            n = n_i;
        end
    end
end
```

### 12.3.9 plotGrid.m (written by Mélodie Hadzic)

```
function plotGrid(h_axes)
prevVal = get(h_axes, 'NextPlot');
set(h_axes,'NextPlot','add');
lim_x = get(h_axes, 'XLim');
lim_y = get(h_axes, 'YLim');
ticks_x = get(h_axes, 'XTick');
ticks_y = get(h_axes, 'YTick');
y_axisY = [ticks_y' ticks_y'];
x_axisY = [lim_x(1) lim_x(2)];
y_axisX = [lim_y(1) lim_y(2)];
x_axisX = [ticks_x' ticks_x'];
for i = 1:size(ticks_x,2)
    plot(h_axes, x_axisX(i,:), y_axisX, '-w');
end
for i = 1:size(ticks_y,2)
    plot(h_axes, x_axisY, y_axisY(i,:), '-w');
end
h_fig = get(h_axes, 'Parent');
set(h_axes,'NextPlot', prevVal);
set(h_axes, 'ButtonDownFcn', {@axes_2Dplot_ButtonDownFcn, h_fig});
```

### 12.3.10 setContPan.m (written by Mélodie Hadzic)

```
function setContPan(str, state, h_fig)
% Update an acion in the control panel
h = guidata(h_fig);
switch state
     case ’error’
          colBg = [1 0.5 0.5];
     case ’success’
          colBg = [0.5 1 0.5];
     case ’process’
          colBg = [1 1 0.5];
     case ’warning’
          colBg = [1 0.75 0.5];
     otherwise
          colBg = [0 0 0];
end
if ~iscell(str)
     str = {str};
end
str = textwrap(h.edit_contPan,str);
set(h.edit_contPan, ’String’, str, ’BackgroundColor’, colBg);
drawnow;
```

## 12.4 Acknowledgements

First and foremost, I thank Prof. Dr. Roland Sigel for shouldering the burden of supervising this thesis, as well as Prof. Dr. David Rueda (Imperial College London, UK) and Prof. Dr. Ben Schuler (University of Zurich, Switzerland) for acting as scientific advisors. Your input indeed helped driving this project to a successful conclusion.

I also thank my colleagues from the Sigel-Freisinger laboratory, in particular Dr. Danny Kowerko who was strongly involved in the research described in this work. Thank you introducing me to the principles and the theory of single-molecule spectroscopy! Your countless insightful remarks and your "innate fussiness" about consistency considerably improved this work. Special thanks are extended to Dr. Richard Börner, Dr. Lucia Cardo, Dr. Daniela Donghi, Alessia Dürst, Erica Fiorini, Prof. Dr. Eva Freisinger, Mélodie Hadzic, Mokrane Khier, Dr. Daniela Kruschel, Dr. Jens Löbus, Dr. Igor Oleinich, Susann Paulus, Dr. Maria Pechlaner, Dr. Joachim Schnabl, Miriam Skilandat, and for their support and manifold contributions to this project. I further want to acknowledge Prof. Dr. Wolf Blanckenhorn (University of Zurich, Switzerland), Dr. Katharina Gapp (Swiss Federal Institute of Technology, Zurich, Switzerland), Prof. Dr. Christian Hübner (University of Lübeck, Germany), Prof. Dr. Achilles C. Kapanidis (University of Oxford, UK), Pramodha Liyanage (Wayne State University, Detroit, USA), Dr. Dylan Muir (University of Basel, Switzerland), Dr. Oliver Otto (Technical University of Dresden, Germany) for scientific input. I also Dr. Richard Börner, Erica Fiorini, Dr. Silke Johannsen and Dr. Danny Kowerko, who proofread this "beast" along with my sister Antje (Wilhelm Büchner University of Applied Sciences, Pfungstadt, Germany).

Last and not least, I want to acknowledge those who kept me on track during my PhD: Eloïse, my parents Norbert and Rosemarie, my brother Torben, my sister Antje, my 34-F (in particular 34-F-78) friends and my friends outside of work. Financial support was granted by the European Research Council (MIRNA 259092), the University of Zurich (Forschungskredit 57010302), the graduate school of chemical and molecular sciences CMSZH and Microsynth AG (CMSZH travel award) and is gratefully acknowledged.

## 12.5 Curriculum Vitae

# Sebastian L.B. König

Diplom Biotechnologe (EUCOR)

Ingénieur en biotechnologie (ESBS)

M.sc Ingénerie et Sciences de l'environnement (EOST)

## EDUCATION

| | |
|---|---|
| **10/10 - present** | **PhD: Institute of Inorganic Chemistry, University of Zurich, CH**<br>Bioinorganic chemistry: “Studying the influence of metal ions on 5’ splice site formation in the group II intron *Sc*.ai5$\gamma$ by single-molecule fluorescence”<br>*Supervision: Prof. Dr. Roland Sigel* |
| **01/10 - 07/10** | **Masters project: Dept. of Physics, University of Cambridge, UK**<br>Biophysics and organic chemistry: “G-quadruplexes and *i*-motifs - influence on duplex stability and synthesis of a new G-quadruplex-binding ligand”<br>*Supervision: Dr Julian Huppert (MP)* |
| **09/08 - 09/10** | **Undergrad.: Ecole et Observatoire des Sciences de la Terre (EOST), FR**<br>M.Sc. programme in environmental science in cooperation with ESBS |
| **09/07 - 09/10** | **Undergrad.: Ecole Sup. de Biotechnologie de Strasbourg (ESBS), FR**<br>Trinational M.Sc. programme in biotechnology (FR, GER, CH) |
| **10/05 - 10/07** | **Undergraduate: University of Karlsruhe, GER**<br>Vordiplom in chemistry |
| **08/95 - 04/04** | **High school: Gymnasium am Kaiserdom, Speyer, GER** |

## WORK EXPERIENCE

| | |
|---|---|
| **07/09 - 08/09** | **Intern: Experimental Toxicology, BASF SE, Ludwigshafen, GER**<br>Validation of a std. operational procedure to determine FMO activity |
| **07/08 - 08/08** | **Intern: Seed Conservation Dept., Kew Gardens, London, UK**<br>Tests of a method to measure dehydroascorbic acid plant material. |
| **05/06 - 12/07** | **Manager: AWIKA GmbH, Karlsruhe, GER**<br>Strategy, HR, controlling, marketing of a personnel service company |
| **07/04 - 06/05** | **Gap year: Emmaus Koenigshoffen, Strasbourg, FR**<br>Taking care of elderly people |

## TEACHING

| | |
|---|---|
| **07/11 - 09/11** | **Supervision: summer internship**<br>Folding of RNA G-quadruplexes on the single-molecule level<br>*Student:* Alessia Dürst, University of Zurich |
| **09/11 - 12/11** | **Practical course: “Chemistry for Biologists”**<br>Responsible for 40 students, 12 h of teaching/week, correcting assignments. |
| **01/12 - 07/12** | **Supervision: masters project**<br>d3’EBS1*-IBS1* interaction in the presence of various metal ions.<br>*Student:* Mokrane Khier, University of Orléans, France |

## RELEVANT SKILLS

| | |
|---|---|
| **Programming** | Matlab, HTML, basic knowledge in Perl and PhP |
| **Software** | MS-Windows, Linux (basic knowledge), MS-Office, MS-Access, InDesign, Photoshop, Chemdraw, Latex and Dreamweaver (basic knowledge) |

## LANGUAGES

| | |
|---|---|
| **German** | Native |
| **English** | Fluent (TOEIC: 980/990, TOEFL: 113/120) |
| **French** | Fluent (DALF C2) |
| **Spanish** | Basic knowledge (3 years) |
| **Arabic** | Basic knowledge (1 year) |
| **Latin** | Intermediate proficiency certificate (6 years) |

## FELLOWSHIPS

| | |
|---|---|
| **07/11 - 04/13** | Forschungskredit: full PhD scholarship, awarded by the University of Zurich |
| **09/12** | SCS Fall Meeting: 2nd prize for the best oral presentation |
| **03/11, 08/12** | CMSZH: full reimbursement of travel and conference participation fees |
| **08/11** | BASF SE: full reimbursement of travel and course participation fees |
| **01/10 - 05/10** | DAAD: German Academic Exchange Service |
| **01/08 - 12/08** | EUCOR: Scholarship of the Upper Rhine Universities |
| **10/07 - 09/08** | Erasmus: Scholarship of the European Union |

**REFEREES**

**Dr Julian Huppert (MP)**
Cavendish Lab
University of Cambridge
JJ Thomson Avenue
Cambridge CB3 0HE, UK
Tel: +44 1223 304421
*jlh29@cam.ac.uk*

**Prof. Dr. Roland Sigel**
Department of Chemistry
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich
Tel: +41 44 635 4652
*roland.sigel@chem.uzh.ch*

## 12.6 Publications and scientific contributions at conferences

Kowerko D., **König S.L.B.**, Böttcher, A., Jelezarov, I., Sigel, R.K.O. Regression-based van't Hoff analysis of thermal melting experiments. Manuscript in preparation.

Kowerko D., Hadzic M., **König S.L.B.**, Börner, R., Heidernätsch, M., Sigel, R.K.O. Multifunctional analysis software for heterogeneous single molecule FRET data. Manuscript in preparation.

**König S.L.B.***, Kowerko D.*, Khier M., Sigel, R.K.O. Cation-promoted RNA tertiary structure formation dissected by single-molecule fluorescence. Submitted to *Nat. Chem.*

Kowerko D.*, **König S.L.B.***, Skilandat M., Kruschel D., Cardo L., Sigel, R.K.O. Metal ion induced kinetic heterogeneity of the intron-exon recognition in single group II introns. Submitted to *Proc. Natl. Acad. Sci. USA*

**König S.L.B.**, Hadzic M., Fiorini E., Börner R., Kowerko D., Blanckenhorn W.U., Sigel R.K.O. (2013). Bootstrap-based analysis of single-molecule FRET data. *PLoS ONE*, **8**, e84157

**König S.L.B.**, Kowerko D., Sigel, R.K.O. (2013). Kinetic subpopulations detected by single-molecule spectroscopy - fundamental property of functional nucleic acids or experimental artefact? *CHIMIA*, **67**, 240-243

**König S.L.B.**, Liyanage P., Sigel, R.K.O., Rueda, D. (2013). Helicase-mediated changes in RNA structure at the single-molecule level (Review). *RNA Biol.* **10**, 132-147

**König S.L.B.**, Huppert J.L., Sigel, R.K.O., Evans A.C.E. (2013). Distance-dependent duplex DNA destabilisation proximal to G-quadruplex/*i*-motif sequences. *Nucleic Acids Res.* **41**, 7453-7461

**Koenig S.L.B.**, Evans A.C.E., Huppert J.L. (2010). Seven essential questions about G-quadruplexes (Review). *Biomol. Concepts* **1(2)**, 197-213

* Equal contributions.

| | |
|---|---|
| **09/13** | Fall meeting of the Swiss Chemical Society, Zurich, Switzerland<br>Poster: *Dissecting cation-dependent RNA secondary structure formation by single-molecule fluorescence* |
| **05/13** | Chemistry of Metals in Biological Systems, Louvain-la-Neuve, Belgium<br>Poster: *Cation-dependent formation of RNA structure revealed by single-molecule fluorescence* |
| **02/13** | 5th European course on time-resolved microscopy, Berlin, Germany |
| **09/12** | Fall meeting of the Swiss Chemical Society, Zurich, Switzerland<br>Talk: *Single-molecule studies of RNA-metal ion interaction* |
| **08/12** | Single-molecule approaches to biology, West Dover, VT, USA<br>Poster: *Metal ion dependent formation of 5' splice site formation revealed by single-molecule FRET* |
| **08/11** | BASF International Summer Course for Scientists, Ludwigshafen, Germany<br>Poster: *Single-molecule studies of group II intron folding* |
| **03/11** | FRET in Life Sciences, Göttingen, Germany |